TABLE 64. N, A/G/T, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
GGT	YES	GLYCINE	4	NONPOLAR (NPL)	21
GGC	YES				
GGA	YES				
GGG	YES				
		ALANINE	0		
GTT	YES	VALINE	4		
GTC	YES				
GTA	YES				
GTG	YES				
TTA	YES	LEUCINE	6		
TTG	YES				
CTT	YES				
CTC	YES				
CTA	YES				
CTG	YES				
ATT	YES	ISOLEUCINE	3		
ATC	YES				
ATA	YES				
ATG	YES	METHIONINE	1		
TTT	YES	PHENYLALANINE	2		
TTC	YES				
TGG	YES	TRYPTOPHAN	1		
		PROLINE	0		
AGT	YES	SERINE	2	POLAR NONIONIZABLE (POL)	10
AGC	YES				
TGT	YES	CYSTEINE	2		
TGC	YES				
AAT	YES	ASPARAGINE	2		
AAC	YES				
CAA	YES	GLUTAMINE	2		
CAG	YES				
TAT	YES	TYROSINE	2		
TAC	YES				
		THREONINE	0		
GAT	YES	ASPARTIC ACID	2	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	4
GAC	YES				
GAA	YES	GLUTAMIC ACID	2		
GAG	YES				
AAA	YES	LYSINE	2	IONIZABLE: BASIC POSITIVE CHARGE (POS)	10
AAG	YES				
CGT	YES	ARGININE	6		
CGC	YES				
CGA	YES				
CGG	YES				
AGA	YES				
AGG	YES				
CAT	YES	HISTIDINE	2		
CAC	YES				
TAA	YES	STOP CODON	3	STOP SIGNAL (STP)	3
TAG	YES				
TGA	YES				
TOTAL		48	17 Amino Acids Are Represented	NPL: POL: NEG: POS: STP 21: 10: 4: 10: 3	

TABLE 65. N, C/G/T, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
GGT	YES	GLYCINE	4	NONPOLAR (NPL)	29
GGC	YES				
GGA	YES				
GGG	YES				
GCT	YES	ALANINE	4		
GCC	YES				
GCA	YES				
GCG	YES				
GTT	YES	VALINE	4		
GTC	YES				
GTA	YES				
GTG	YES				
TTA	YES	LEUCINE	6		
TTG	YES				
CTT	YES				
CTC	YES				
CTA	YES				
CTG	YES				
ATT	YES	ISOLEUCINE	3		
ATC	YES				
ATA	YES				
ATG	YES	METHIONINE	1		
TTT	YES	PHENYLALANINE	2		
TTC	YES				
TGG	YES	TRYPTOPHAN	1		
CCT	YES	PROLINE	4		
CCC	YES				
CCA	YES				
CCG	YES				
TCT	YES	SERINE	6	POLAR NONIONIZABLE (POL)	12
TCC	YES				
TCA	YES				
TCG	YES				
AGT	YES				
AGC	YES				
TGT	YES	CYSTEINE	2		
TGC	YES				
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
ACT	YES	THREONINE	4		
ACC	YES				
ACA	YES				
ACG	YES				
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	6
CGT	YES	ARGININE	6		
CGC	YES				
CGA	YES				
CGG	YES				
AGA	YES				
AGG	YES				
		HISTIDINE	0		
TGA	YES	STOP CODON	1	STOP SIGNAL (STP)	1
TOTAL		48	13 Amino Acids Are Represented	NPL: POL: NEG: POS: STP 29: 12: 0: 6: 1	

TABLE 66. C, C, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	4
		ALANINE	0		
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
CCT	YES	PROLINE	4		
CCC	YES				
CCA	YES				
CCG	YES				
		SERINE	0	POLAR NONIONIZABLE (POL)	0
		CYSTEINE	0		
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
		THREONINE	0		
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
		ARGININE	0		
		HISTIDINE	0		
		STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL	4	1 Amino Acid Is Represented		NPL: POL: NEG: POS: STP 4: 0: 0: 0: 0	

TABLE 67. G, G, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
GGT	YES	GLYCINE	4	NONPOLAR (NPL)	4
GGC	YES				
GGA	YES				
GGG	YES				
		ALANINE	0		
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
		PROLINE	0		
		SERINE	0	POLAR NONIONIZABLE (POL)	0
		CYSTEINE	0		
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
		THREONINE	0		
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
		ARGININE	0		
		HISTIDINE	0		
		STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL	0 4	1 Amino Acid Is Represented		NPL: POL: NEG: POS: STP 4: 0: 0: 0: 0	

TABLE 68. G, C, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	4
GCT	YES	ALANINE	4		
GCC	YES				
GCA	YES				
GCG	YES				
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
		PROLINE	0		
		SERINE	0	POLAR NONIONIZABLE (POL)	0
		CYSTEINE	0		
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
		THREONINE	0		
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
		ARGININE	0		
		HISTIDINE	0		
		STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL	4	1 Amino Acid Is Represented		NPL: POL: NEG: POS: STP 4: 0: 0: 0: 0	

TABLE 69. G, T, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	4
		ALANINE	0		
GTT	YES	VALINE	4		
GTC	YES				
GTA	YES				
GTG	YES				
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
		PROLINE	0		
		SERINE	0	POLAR NONIONIZABLE (POL)	0
		CYSTEINE	0		
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
		THREONINE	0		
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
		ARGININE	0		
		HISTIDINE	0		
		STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL	4	1 Amino Acid Is Represented		NPL: POL: NEG: POS: STP 4: 0: 0: 0: 0	

TABLE 70. C, G, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	0
		ALANINE	0		
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
		PROLINE	0		
		SERINE	0	POLAR NONIONIZABLE (POL)	0
		CYSTEINE	0		
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
		THREONINE	0		
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	4
CGT	YES	ARGININE	4		
CGC	YES				
CGA	YES				
CGG	YES				
		HISTIDINE	0		
		STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL	4	1 Amino Acid Is Represented		NPL: POL: NEG: POS: STP 0: 0: 0: 4: 0	

TABLE 71. C, T, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	4
		ALANINE	0		
		VALINE	0		
CTT	YES	LEUCINE	4		
CTC	YES				
CTA	YES				
CTG	YES				
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
		PROLINE	0		
		SERINE	0	POLAR NONIONIZABLE (POL)	0
		CYSTEINE	0		
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
		THREONINE	0		
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
		ARGININE	0		
		HISTIDINE	0		
		STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL	4	1 Amino Acid Is Represented		NPL: POL: NEG: POS: STP 4: 0: 0: 0: 0	

TABLE 72. T, C, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	0
		ALANINE	0		
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
		PROLINE	0		
TCT	YES	SERINE	4	POLAR NONIONIZABLE (POL)	4
TCC	YES				
TCA	YES				
TCG	YES				
		CYSTEINE	0		
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
		THREONINE	0		
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
		ARGININE	0		
		HISTIDINE	0		
		STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL	4	1 Amino Acid Is Represented		NPL: POL: NEG: POS: STP 0: 4: 0: 0: 0	

5

TABLE 73. A, C, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	0
		ALANINE	0		
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
		PROLINE	0		
		SERINE	0	POLAR NONIONIZABLE (POL)	4
		CYSTEINE	0		
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
ACT	YES	THREONINE	4		
ACC	YES				
ACA	YES				
ACG	YES				
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
		ARGININE	0		
		HISTIDINE	0		
		STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL	4	1 Amino Acid Is Represented		NPL: POL: NEG: POS: STP 0: 4: 0: 0: 0	

TABLE 74. G, A, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	0
		ALANINE	0		
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
		PROLINE	0		
		SERINE	0	POLAR NONIONIZABLE (POL)	0
		CYSTEINE	0		
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
		THREONINE	0		
GAT	YES	ASPARTIC ACID	2	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	4
GAC	YES				
GAA	YES	GLUTAMIC ACID	2		
GAG	YES				
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
		ARGININE	0		
		HISTIDINE	0		
		STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL	4	2 Amino Acids Are Represented		NPL: POL: NEG: POS: STP 0: 0: 4: 0: 0	

TABLE 75. A, T, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	4
		ALANINE	0		
		VALINE	0		
		LEUCINE	0		
ATT	YES	ISOLEUCINE	3		
ATC	YES				
ATA	YES				
ATG	YES	METHIONINE	1		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
		PROLINE	0		
		SERINE	0	POLAR NONIONIZABLE (POL)	0
		CYSTEINE	0		
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
		THREONINE	0		
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
		ARGININE	0		
		HISTIDINE	0		
		STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL	4	2 Amino Acids Are Represented		NPL: POL: NEG: POS: STP 4: 0: 0: 0: 0	

TABLE 76. C, A, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	0
		ALANINE	0		
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
		PROLINE	0		
		SERINE	0	POLAR NONIONIZABLE (POL)	2
		CYSTEINE	0		
		ASPARAGINE	0		
CAA	YES	GLUTAMINE	2		
CAG	YES				
		TYROSINE	0		
		THREONINE	0		
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	2
		ARGININE	0		
CAT	YES	HISTIDINE	2		
CAC	YES				
		STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL	4	2 Amino Acids Are Represented		NPL: POL: NEG: POS: STP 0: 2: 0: 2: 0	

TABLE 77. T, T, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	4
		ALANINE	0		
		VALINE	0		
TTA	YES	LEUCINE	2		
TTG	YES				
		ISOLEUCINE	0		
		METHIONINE	0		
TTT	YES	PHENYLALANINE	2		
TTC	YES				
		TRYPTOPHAN	0		
		PROLINE	0		
		SERINE	0	POLAR NONIONIZABLE (POL)	0
		CYSTEINE	0		
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
		THREONINE	0		
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
		ARGININE	0		
		HISTIDINE	0		
		STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL	4	2 Amino Acids Are Represented		NPL: POL: NEG: POS: STP 4: 0: 0: 0: 0	

TABLE 78. A, A, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	0
		ALANINE	0		
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
		PROLINE	0		
		SERINE	0	POLAR NONIONIZABLE (POL)	2
		CYSTEINE	0		
AAT	YES	ASPARAGINE	2		
AAC	YES				
		GLUTAMINE	0		
		TYROSINE	0		
		THREONINE	0		
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
AAA	YES	LYSINE	2	IONIZABLE: BASIC POSITIVE CHARGE (POS)	2
AAG	YES				
		ARGININE	0		
		HISTIDINE	0		
		STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL	4	2 Amino Acids Are Represented		NPL: POL: NEG: POS: STP 0: 2: 0: 2: 0	

5

TABLE 79. T, A, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	0
		ALANINE	0		
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
		PROLINE	0		
		SERINE	0	POLAR NONIONIZABLE (POL)	2
		CYSTEINE	0		
		ASPARAGINE	0		
		GLUTAMINE	0		
TAT	YES	TYROSINE	2		
TAC	YES				
		THREONINE	0		
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
		ARGININE	0		
		HISTIDINE	0		
TAA	YES	STOP CODON	2	STOP SIGNAL (STP)	2
TAG	YES				
TOTAL	4	1 Amino Acid Is Represented		NPL: POL: NEG: POS: STP 0: 2: 0: 0: 2	

TABLE 80. T, G, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	1
		ALANINE	0		
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
TGG	YES	TRYPTOPHAN	1		
		PROLINE	0		
		SERINE	0	POLAR NONIONIZABLE (POL)	2
TGT	YES	CYSTEINE	2		
TGC	YES				
		TYROSINE	0		
		THREONINE	0		
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
		ARGININE	0		
		HISTIDINE	0		
TGA	YES	STOP CODON	1	STOP SIGNAL (STP)	1
TOTAL	4	2 Amino Acids Are Represented		NPL: POL: NEG: POS: STP 1: 2: 0: 0: 1	

5

TABLE 81. A, G, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	0
		ALANINE	0		
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
		PROLINE	0		
AGT	YES	SERINE	2	POLAR NONIONIZABLE (POL)	2
AGC	YES				
		CYSTEINE	0		
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
		THREONINE	0		
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	2
AGA	YES	ARGININE	2		
AGG	YES				
		HISTIDINE	0		
		STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL	4	2 Amino Acids Are Represented		NPL: POL: NEG: POS: STP 0: 2: 0: 2: 0	

TABLE 82. G/C, G, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
GGT	YES	GLYCINE	4	NONPOLAR (NPL)	4
GGC	YES				
GGA	YES				
GGG	YES				
		ALANINE	0		
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
		PROLINE	0		
		SERINE	0	POLAR NONIONIZABLE (POL)	0
		CYSTEINE	0		
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
		THREONINE	0		
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	4
CGT	YES	ARGININE	4		
CGC	YES				
CGA	YES				
CGG	YES				
		HISTIDINE	0		
		STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL	8	2 Amino Acids Are Represented		NPL: 4: POL: 0: NEG: 0: POS: 4: STP 0	

TABLE 83. G/C, C, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	8
GCT	YES	ALANINE	4		
GCC	YES				
GCA	YES				
GCG	YES				
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
CCT	YES	PROLINE	4		
CCC	YES				
CCA	YES				
CCG	YES				
		SERINE	0	POLAR NONIONIZABLE (POL)	0
		CYSTEINE	0		
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
		THREONINE	0		
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
		ARGININE	0		
		HISTIDINE	0		
		STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL	8	2 Amino Acids Are Represented		NPL: 8: POL: 0: NEG: 0: POS: 0: STP 0	

TABLE 84. G/C, A, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	0
		ALANINE	0		
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
		PROLINE	0		
		SERINE	0	POLAR NONIONIZABLE (POL)	2
		CYSTEINE	0		
		ASPARAGINE	0		
CAA	YES	GLUTAMINE	2		
CAG	YES				
		TYROSINE	0		
		THREONINE	0		
GAT	YES	ASPARTIC ACID	2	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	4
GAC	YES				
GAA	YES	GLUTAMIC ACID	2		
GAG	YES				
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	2
		ARGININE	0		
CAT	YES	HISTIDINE	2		
CAC	YES				
		STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL	8	4 Amino Acids Are Represented		NPL: 0: POL: 2: NEG: 4: POS: 2: STP 0	

5

TABLE 85. G/C, T, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	8
		ALANINE	0		
GTT	YES	VALINE	4		
GTC	YES				
GTA	YES				
GTG	YES				
CTT	YES	LEUCINE	4		
CTC	YES				
CTA	YES				
CTG	YES				
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
		PROLINE	0		
		SERINE	0	POLAR NONIONIZABLE (POL)	0
		CYSTEINE	0		
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
		THREONINE	0		
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
		ARGININE	0		
		HISTIDINE	0		
		STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL	8	2 Amino Acids Are Represented		NPL: 8: POL: 0: NEG: 0: POS: 0: STP 0	

Methods of administering packaged nucleic acids in mammals for transduction of cells *in vivo*

Genetic vaccine vectors (e.g., adenoviruses, liposomes, papillomaviruses, retroviruses, etc.) can be administered directly to the mammal for transduction of cells *in vivo*. The genetic vaccines obtained using the methods of the invention can be formulated as pharmaceutical compositions for administration in any suitable manner, including parenteral (e.g., subcutaneous, intramuscular, intradermal, or intravenous), topical, oral, rectal, intrathecal, buccal (e.g., sublingual), or local administration, such as by aerosol or transdermally, for prophylactic and/or therapeutic treatment. Pretreatment of skin, for example, by use of hair-removing agents, may be useful in transdermal delivery. Suitable methods of administering such packaged nucleic acids are available and well known to those of skill in the art, and, although more than one route can be used to administer a particular composition, a particular route can often provide a more immediate and more effective reaction than another route.

Pharmaceutically acceptable carriers are determined in part by the particular composition being administered, as well as by the particular method used to administer the composition. Accordingly, there is a wide variety of suitable formulations of pharmaceutical compositions of the present invention. A variety of aqueous carriers can be used, e.g., buffered saline and the like. These solutions are sterile and generally free of undesirable matter. These compositions may be sterilized by conventional, well known sterilization techniques. The compositions may contain pharmaceutically acceptable auxiliary substances as required to approximate physiological conditions such as pH adjusting and buffering agents, toxicity adjusting agents and the like, for example, sodium acetate, sodium chloride, potassium chloride, calcium chloride, sodium lactate and the like. The concentration of genetic vaccine vector in these formulations can vary widely, and will be selected primarily based on fluid volumes, viscosities, body weight and the like in accordance with the particular mode of administration selected and the patient's needs.

Formulations suitable for oral administration can consist of (a) liquid solutions, such as an effective amount of the packaged nucleic acid suspended in diluents, such as water, saline or PEG 400; (b) capsules, sachets or tablets, each containing a predetermined amount of the active ingredient, as liquids, solids, granules or gelatin; (c) suspensions in an appropriate liquid; and (d) suitable emulsions. Tablet fonns can include one or more of

lactose, sucrose, mannitol, sorbitol, calcium phosphates, corn starch, potato starch, tragacanth, microcrystalline cellulose, acacia, gelatin, colloidal silicon dioxide, croscannellose sodium, talc, magnesium stearate, stearic acid, and other excipients, colorants, fillers, binders, diluents, buffering agents, moistening agents, preservatives, flavoring agents, dyes, disintegrating agents, and pharmaceutically compatible carriers.

Lozenge forms can comprise the active ingredient in a flavor, usually sucrose and acacia or tragacanth, as well as pastilles comprising the active ingredient in an inert base, such as gelatin and glycerin or sucrose and acacia emulsions, gels, and the like containing, in addition to the active ingredient, carriers known in the art. It is recognized that the genetic vaccines, when administered orally, must be protected from digestion. This is typically accomplished either by complexing the vaccine vector with a composition to render it resistant to acidic and enzymatic hydrolysis or by packaging the vector in an appropriately resistant carrier such as a liposome. Means of protecting vectors from digestion are well known in the art. The pharmaceutical compositions can be encapsulated, e. g., in liposomes, or in a formulation that provides for slow release of the active ingredient.

The packaged nucleic acids, alone or in combination with other suitable components, can be made into aerosol formulations (e.g., they can be "nebulized") to be administered via inhalation. Aerosol formulations can be placed into pressurized acceptable propellants, such as dichlorodifluoromethane, propane, nitrogen, and the like. Suitable formulations for rectal administration include, for example, suppositories, which consist of the packaged nucleic acid with a suppository base. Suitable suppository bases include natural or synthetic triglycerides or paraffin hydrocarbons. In addition, it is also possible to use gelatin rectal capsules which consist of a combination of the packaged nucleic acid with a base, including, for example, liquid triglycerides, polyethylene glycols, and paraffin hydrocarbons.

Formulations suitable for parenteral. administration, such as, for example, by intraarticular (in the joints), intravenous, intramuscular, intradermal, intraperitoneal, and subcutaneous routes, include aqueous and non-aqueous, isotonic sterile injection solutions, which can contain antioxidants, buffers, bacteriostats, and solutes that render the formulation isotonic with the blood of the intended recipient, and aqueous and non- aqueous sterile suspensions that can include suspending agents, solubilizers, thickening agents, stabilizers, and preservatives. In the practice of this invention, compositions can be administered, for

example, by intravenous infusion, orally, topically, intraperitoneally, intravesically or intrathecally.

Parenteral administration and intravenous administration are the preferred methods of administration

The formulations of packaged nucleic acid can be presented in unit-dose or multi-dose sealed containers, such as ampoules and vials. Injection solutions and suspensions can be prepared from sterile powders, granules, and tablets of the kind previously described. Cells transduced by the packaged nucleic acid can also be administered intravenously or parenterally.

Dose size

The dose administered to a patient, in the context of the present invention should be sufficient to effect a beneficial therapeutic response in the patient over time. The dose will be determined by the efficacy of the particular vector employed and the condition of the patient, as well as the body weight or vascular surface area of the patient to be treated. The size of the dose also will be determined by the existence, nature, and extent of any adverse side-effects that accompany the administration of a particular vector, or transduced cell type in a particular patient.

In determining the effective amount of the vector to be administered in the treatment or prophylaxis of an infection or other condition, the physician evaluates vector toxicities, progression of the disease, and the production of anti- vector antibodies, if any. In general, the dose equivalent of a naked nucleic acid from a vector is from about 1μg to 1mg for a typical 70 kilogram patient, and doses of vectors used to deliver the nucleic acid are calculated to yield an equivalent amount of therapeutic nucleic acid. Administration can be accomplished via single or divided doses.

In therapeutic applications, compositions are administered to a patient suffering from a disease (e.g., an infectious disease or autoimmune disorder) in an amount sufficient to cure or at least partially arrest the disease and its complications. An amount adequate to accomplish this is defined as a "therapeutically effective dose." Amounts effective for this use will depend upon the severity of the disease and the general state of the patient's health. Single or multiple administrations of the compositions may be administered depending on the dosage and frequency as required and tolerated by the patient. In any event, the composition

should provide a sufficient quantity of the proteins of this invention to effectively treat the patient.

In prophylactic applications, compositions are administered to a human or other mammal to induce an immune response that can help protect against the establishment of an infectious disease or other condition.

Ability to determine toxicity therapeutic efficacy

The toxicity and therapeutic efficacy of the genetic vaccine vectors provided by the invention are determined using standard pharmaceutical procedures in cell cultures or experimental animals. One can determine the LD_{50} (the dose lethal to 50% of the population) and the ED_{50} (the dose therapeutically effective in 50% of the population) using procedures presented herein and those otherwise known to those of skill in the art.

More on dosage

A typical pharmaceutical composition for intravenous administration would be about 0.1 to 10 mg per patient per day. Dosages from 0.1 up to about 100 mg per patient per day may be used, particularly when the drug is administered to a secluded site and not into the blood stream, such as into a body cavity or into a lumen of an organ. Substantially higher dosages are possible in topical administration. Actual methods for preparing parenterally administrable compositions will be known or apparent to those skilled in the art and are described in more detail in such publications as Remington's Pharmaceutical Science, 15th ed., Mack Publishing Company, Easton, Pennsylvania (1980).

Packaging/dispenser devices

The genetic vaccines obtained using the methods of the invention (e.g. the multivalent antigenic polypeptides of the invention, and genetic vaccines that express the polypeptides) can be packaged in packs, dispenser devices, and kits for administering genetic vaccines to a mammal. For example, packs or dispenser devices that contain one or more unit dosage forms are provided. Typically, instructions for administration of the compounds will be provided with the packaging, along with a suitable indication on the label that the compound is suitable for treatment of an indicated condition. For example, the label may state that the active compound within the packaging is useful for treating a particular infectious disease, autoimmune disorder, tumor, or for preventing or treating other diseases or conditions that are mediated by, or potentially susceptible to, a mammalian immune response.

USES OF GENETIC VACCINES

Genetic vaccines which include optimized vector modules and other reagents provided by the invention are useful for treating many diseases and other conditions that are either mediated by a mammalian immune system or are susceptible to treatment by an appropriate immune response. Representative examples of these diseases & antigens appropriate for each are listed below, described herein, or incorporated by reference.

Substrates for evolution of optimized recombinant antigens

The invention provides methods of obtaining experimentally generated polynucleotides that encode antigens that exhibit improved ability to induce an immune response to a pathogenic agent. The methods are applicable to a wide range of pathogenic agents, including potential biological warfare agents and other organisms and polypeptides that can cause disease and toxicity in humans and other animals. The following examples are merely illustrative, and not limiting.

Infectious Diseases

Genetic vaccine vectors obtained according to the methods of the invention are useful in both prophylaxis and therapy of infectious diseases, including those caused by any bacterial, fungal, viral, or other pathogens of mammals. In some embodiments, protection is conferred by use of a genetic vaccine vector that will express an antigen (either or both of a humoral antigen or a T cell antigen) of the pathogen of interest. In preferred embodiments, the antigen is evolved using the methods of the invention in order to obtain optimized antigens as described herein. The vector induces an immune response against the antigen. One or several antigens or antigen fragments can be included in one genetic vaccine delivery vehicle. Examples of pathogens and corresponding polypeptides from which an antigen can be obtained include, but are not limited to, HIV (gp120, gp160), hepatitis B, C, D, E (surface antigen), rabies (glycoprotein), Schistosoma mansoni (calpain; Jankovic (1996) J Immunol. 157: 806-14). Other pathogen infections that are treatable using genetic vaccine vectors include, for example, herpes zoster, herpes simplex -1 and -2, tuberculosis (including chronic, drug-resistant), lyme disease (*Borrelia burgorferii*), syphilis, parvovirus, rabies, human papillomavirus, and the like.

Bacterial Pathogens And Toxins

In some embodiments, the methods of the invention are applied to bacterial pathogens, as well as to toxins produced by bacteria and other organisms. One can use the methods to obtain experimentally generated polypeptides that can induce an immune response against the pathogen, as well as recombinant toxins that are less toxic than native toxin polypeptides. Often, the polynucleotides of interest encode polypeptides that are present on the surface of the pathogenic organism. Among the pathogens for which the methods of the invention are useful for producing protective immunogenic experimentally generated polypeptides are the Yersinia species.

Yersinia pestis, the causative agent of plague, is one of the most virulent bacteria known with LD_{50} values in mouse of less than 10 bacteria. The pneumonic form of the disease is readily spread between humans by aerosol or infectious droplets and can be lethal within days. A particularly preferred target for obtaining a experimentally generated polypeptide that can protect against Yersinia infection is the V antigen, which is a 37 kDa virulence factor, induces protective immune responses and is currently being evaluated as a subunit vaccine (Brubaker (1991) Current Investigations of the Microbiology of Yersinae, 12: 127). The V-antigen alone is not toxic, but Y pestis isolates that lack the V-antigen are avirulent. The Yersinia V- antigen has been successfully produced in E. coli by several groups (Leary et al. (1995) Infect. Immun. 3: 2854). Antibodies that recognize the V-antigen can provide passive protection against homologous strains, but not against heterologous strains. Similarly, immunization with purified V antigen protects against only homologous strains. To obtain cross-protective recombinant V antigen, in a preferred embodiment, V antigen genes from various Yersinia species are subjected to polynucleotide reassembly (optionally in combination with other directed evolution methods described herein). The genes encoding the V antigen from Y. pestis, Y. enterocolitica, and Y. pseudotuberculosis, for example, are 92-99% identical at the DNA level, making them ideal for optimization using family reassembly (optionally in combination with other directed evolution methods described herein) according to the methods of the invention. After reassembly (optionally in combination with other directed evolution methods described herein), the library of recombinant nucleic acids is screened and/or selected for those that encode recombinant V antigen polypeptides that can induce an improved immune response and/or have greater cross- protectivity.

Bacillus anthracis, the causative agent of anthrax, is another example of a bacterial target against which the methods of the invention are useful. The anthrax protective antigen (PA) provides protective immune responses in test animals, and antibodies against PA also provide some protection. However, the immunogenicity of PA is relatively poor, so multiple injections are typically required when wild-type PA is used. Co- vaccination with lethal factor (LF) can improve the efficacy of wild-type PA vaccines, but toxicity is a limiting factor. Accordingly the stochastic (e.g. polynucleotide shuffling & interrupted synthesis) and non-stochastic polynucleotide reassembly and antigen library immunization methods of the invention can be used to obtain nontoxic LF. Polynucleotides that encode LF from various B. anthracis strains are subjected to family reassembly (optionally in combination with other directed evolution methods described herein). The resulting library of recombinant LF nucleic acids can then be screened to identify those that encode recombinant LF polypeptides that exhibit reduced toxicity. For example, one can inoculate tissue culture cells with the recombinant LF polypeptides in the presence of PA and select those clones for which the cells survive. If desired, one can then backcross the nontoxic LF polypeptides to retain the immunogenic epitopes of LF. Those that are selected through the first screen can then be subjected to a secondary screen. For example, one can test for the ability of the recombinant nontoxic LF polypeptides to induce an immune response (e.g., CTL or antibody response) in a test animal such as mice. In preferred embodiments, the recombinant nontoxic LF polypeptides are then tested for ability to induce protective immunity in test animals against challenge by different strains of B. anthracis.

The protective antigen (PA) of B. anthracis is also a suitable target for the methods of the invention. PA-encoding nucleic acids from various strains of B. anthracis are subjected to stochastic (e.g. polynucleotide shuffling & interrupted synthesis) and non-stochastic polynucleotide reassembly. One can then screen for proper folding in, for example, E. coli, using polyclonal antibodies. Screening for ability to induce broad- spectrum antibodies in a test animal is also typically used, either alone or in addition to a preliminary screening method. In presently preferred embodiments, those experimentally generated polynucleotides that exhibit the desired properties can be backcrossed so that the immunogenic epitopes are maintained. Finally, the selected recombinants are tested for ability to induce protective immunity against different strains of B. anthracis in a test animal.

The Staphylococcus aureus and Streptococcus toxins are another example of a target polypeptide that can be altered using the methods of the invention. Strains of Stapkvlococcus aureus and group A Streptococci are involved in a range of diseases, including food poisoning, toxic shock syndrome, scarlet fever and various autoimmune disorders. They secrete a variety of toxins, which include at least five cytolytic toxins, a coagulase, and a variety of enterotoxins. The enterotoxins are classified as superantigens in that they crosslink MHC class II molecules with T cell receptors to cause a constitutive T cell activation (Fields et al. (1996) Nature 384: 188). This results in the accumulation of pathogenic levels of cytokines that can lead to multiple organ failure and death. At least thirty related, yet distinct enterotoxins have been sequenced and can be phylogenetically grouped into families. Crystal structures have been obtained for several members alone and in complex with MHC class II molecules. Certain mutations in the MHC class II binding site of the toxins strongly reduce their toxicity and can form the basis of attenuated vaccines (Woody et al. (1997) Vaccine 15: 133). However, a successful immune response to one type of toxin may provide protection against closely related family members, whereas little protection against toxins from the other families is observed. Family reassembly (optionally in combination with other directed evolution methods described herein) of enterotoxin genes from various family members can be used to obtain recombinant toxin molecules that have reduced toxicity and can induce a cross-protective immune response. Experimentally evolved (e.g. by polynucleotide reassembly &/or polynucleotide site-saturation mutagenesis) antigens can also be screened to identify antigens that elicit neutralizing antibodies in an appropriate animal model such as mouse or monkey. Examples of such assays can include ELISA formats in which the elicited antibodies prevent binding of the enterotoxin to the MHC complex and/or T cell receptors on cells or purified forms. These assays can also include formats where the added antibodies would prevent T cells from being cross-linked to appropriate antigen presenting cells.

Cholera is an ancient, potentially lethal disease caused by the bacterium Vibrio cholerae and an effective vaccine for its prevention is still unavailable. Much of the pathogenesis of this disease is caused by the cholera enterotoxin. Ingestion of microgram quantities of cholera toxin can induce severe diarrhea causing loss of tens of liters of fluid.

Cholera toxin is a complex of a single catalytic A subunit with a pentameric ring of identical B subunits. Each subunit is inactive on its own. The B subunits bind to specific ganglioside receptors on the surface of intestinal epithelial cells and trigger the entry of the A

subunit into the cell. The A subunit ADP-ribosylates a regulatory G protein initiating a cascade of events causing a massive, sustained flow of electrolytes and water into the intestinal lumen resulting in extreme diarrhea.

The B subunit of cholera toxin is an attractive vaccine target for a number of reasons. It is a major target of protective antibodies generated during cholera infection and contains the epitopes for antitoxin neutralizing antibodies. It is nontoxic without the A subunit, is orally effective, and stimulates production of a strong IgA- dominated gut mucosal immune response, which is essential in protection against cholera and cholera toxin. The B subunit is also being investigated for use as an adjuvant in other vaccine preparations, and therefore, evolved toxins may provide general improvements for a variety of different vaccines. The heat-labile enterotoxins (LT) from enterotoxigenic E. coli strains are structurally related to cholera toxin and are 75% identical at the DNA sequence level. To obtain optimized recombinant toxin molecules that exhibit reduced toxicity and increased ability to induce an immune response that is protective against V. cholerae and E. coli, the genes that encode the related toxins are subjected to stochastic (e.g. polynucleotide shuffling & interrupted synthesis) and non-stochastic polynucleotide reassembly.

The recombinant toxins are then tested for one or more of a several desirable traits. For example, one can screen for improved cross-reactivity of antibodies raised against the recombinant toxin polypeptides, for lack of toxicity in a cell culture assay, and for ability to induce a protective immune response against the pathogens and/or against the toxins themselves. The experimentally evolved (e.g. by polynucleotide reassembly &/or polynucleotide site-saturation mutagenesis) clones can be selected by phage display and/or screened by phage ELISA and ELISA assays for the presence of epitopes from the different serotypes. Variant proteins with multiple epitopes can then be purified and used to immunize mice or other test animal. The animal serum is then assayed for antibodies to the different B chain subtypes and variants that elicit a broad cross-reactive response will be evaluated further in a virulent challenge model. The E. coli and V. cholerae toxins can also act as adjuvants that are capable of enhancing mucosal immunity and oral delivery of vaccines and proteins.

Accordingly, one can test the library of recombinant toxins for enhancement of the adjuvant activity

Experimentally evolved (e.g. by polynucleotide reassembly &/or polynucleotide site-saturation mutagenesis) antigens can also be screened for improved expression levels and stability of the B chain pentamer, which may be less stable than when in the presence of the A chain in the hexameric complex. Addition of a heat treatment step or denaturing agents such as salts, urea, and/or guanidine hydrochloride can be included prior to ELISA assays to measure yields of correctly folded molecules by appropriate antibodies. It is sometimes desirable to screen for stable monomeric B chain molecules, in an ELISA format, for example, using antibodies that bind monomeric, but not pentameric B chains. Additionally, the ability of experimentally evolved (e.g. by polynucleotide reassembly &/or polynucleotide site-saturation mutagenesis) antigens to elicit neutralizing antibodies in an appropriate animal model such as mouse or monkey can be screened. For example, antibodies that bind to the B chain and prevent its binding to its specific ganglioside receptors on the surface of intestinal epithelial cells may prevent disease. Similarly antibodies that bind to the B chain and prevent its pentamerization or block A chain binding may be useful in preventing disease.

The bacterial antigens that can be improved by stochastic (e.g. polynucleotide shuffling & interrupted synthesis) and non-stochastic polynucleotide reassembly for use as vaccines also include, but are not limited to, Helicobacter pylori antigens CagA and VacA (Blaser (1996) Aliment. Pharmacol. Ther. 1: 73-7; Blaser and Crabtree (1996) Am. J Clin. Pathol. 106: 565-7; Censini et al. (1996) Proc. Nat'l. Acad. Sci. USA 93: 14648-14643).

Other suitable H. pylori antigens include, for example, four immunoreactive proteins of 45-65 kDa as reported by Chatha et al. (1997) Indian J Med. Res. 105: 170- 175 and the H. pylori GroES homologue (HspA) (Kansau et al. (1996) Mol. Microbiol. 22: 1013-1023. Other suitable bacterial antigens include, but are not limited to, the 43-kDa and the fimbrilin (41 kDa) proteins of P. gingivalis (Boutsl et al. (1996) Oral Microbiol. Immunol. 11: 236-241); pneumococcal surface protein A (Briles et al. (1996) Ann. NYAcad. Sci. 797: 118-126); Chlamydia psittaci antigens, 80-90 kDa protein and 110 kDa protein (Buendia et al. (1997) FEMSMicrobiol. Lett. 150: 113-9); the chlainydial exoglycolipid antigen (GLXA) (Whittum-Hudson et al. (1996) Nature Med. 2: 1116-112 1); Chlamlydia pneumoniae species- specific antigens in the molecular weight ranges 92-98, 51-55, 43-46 and 31.5-33 kDa and genus-specific antigens in the ranges 12, 26 and 65-70 kDa (Halme et al. (1997) Scand. J Immunol. 45: 378-84); Neisseria gonorrhoeae (GC) or Escherichia coli phase-variable opacity (Opa) proteins (Chen and Gotschlich (1996) Proc. Nat'l. Acad. Sci. USA 93:

14851-14856), any of the twelve immunodominant proteins of Schistosoma mansoni (ranging in molecular weight from 14 to 208 kDa) as described by Cutts and Wilson (1997) Parasitolog-v 114: 245-55; the 17-kDa protein antigen of Brucella abortus (De Mot et al. (1996) Curr. Microbiol. 33: 26-30); a gene homolog of the 17-kDa protein antigen of the Gram-negative pathogen Brucella abortus identified in the nocardioform actinomycete Rhodococcus sp. N186/21 (De Mot et al. (1996) Curr. Microbiol. 33: 26- 30); the staphylococcal enterotoxins (SEs) (Wood et al. (1997) FEMS Immunol. Med. Microbiol. 17: 1-10), a 42-kDa M. hy,opneunioniae NrdF ribonucleotide reductase R2 protein or 15-kDa subunit protein of M. hyopneumoniae (Fagan et al. (1997) Infect. Immun. 65: 2502-2507), the meningococcal antigen PorA protein (Feavers et al. (1997) Clin. Diagn. Lab. Immunol. 3: 444-50); pneumococcal surface protein A (PspA) (McDaniel et al. (1997) Gene Ther. 4: 375-377); F. tularensis outer membrane protein FopA (Fulop et al. (1996) FEMSImmunol. Med. Microbiol. 13: 245-247); the major outer membrane protein within strains of the genus Actinobacillus (Hartmann et al. (1996) Zentralbl. Bakteriol. 284: 255- 262); p60 or listeriolysin (Hly) antigen of Listeria monocytogenes (Hess et al. (1996) Proc. Nat'l. Acad. Sci. USA 93: 1458-1463); flagellar (G) antigens observed on Salmonella enteritidis and S. pullorum (Holt and Chaubal (1997) J. Clin. Microbiol. 35: 1016-1020); Bacillus anthracis protective antigen (PA) (1vins et al. (1995) Vaccine 13: 1779-1784); Echinococcus granulosus antigen 5 (Jones et al. (1996) Parasitology 113: 213-222); the rol genes of Shigella dvsenteriae I and Escherichia coli K- 12 (Klee et al. (1997) J. Bacteriol. 179: 2421 - 2425); cell surface proteins Rib and alpha of group B streptococcus (Larsson et al. (1996) Infect. Immun. 64: 3518-3523); the 37 kDa secreted polypeptide encoded on the 70 kb virulence plasmid of pathogenic Yersinia spp. (Leary et al. (1995) Contrib. Microbiol. 1mmunol. 13: 216-217 and Roggenkamp et al. (1997) Infect. Immun. 65: 446- 5 1); the OspA (outer surface protein A) of the Lyme disease spirochete Borrelia burgdorferi (Li et al. (1997) Proc. Nat'l. Acad Sci. USA 94: 3584-3589, Padilla et al. (1996) J Infect. Dis. 174: 739-746, and Wallich et al. (1996) Infection 24: 396-397); the Brucella melitensis group 3 antigen gene encoding Omp28 (Lindler et al. (1996) Infect. Immun. 64: 2490-2499); the PAc antigen of Streptococcus mutans (Murakami et al. (1997) Infect. Immun. 65: 794-797); pneumolysin, Pneumococcal neuraminidases, autolysin, hyaluronidase, and the 37 kDa pneumococcal surface adhesin A (Paton et al. (1997) Microb. Drug Resist. 3: 1 -10); 29-32, 41-45, 63-71 x 10(3) MW antigens of Salmonella typhi (Perez et al. (1996) Immunology 89:

262-267); K-antigen as a marker of Klebsiella pneumoniae (Priamukhina and Morozova (1996) Klin. Lab. Diagn. 47-9); nocardial antigens of molecular mass approximately 60, 40, and 15-10 kDa (Prokesova et al. (1996) Int. J Immunopharmacol. 18: 661- 668); Staphylococcus aureus antigen ORF-2 (Rieneck et al. (1997) Biochim Biophys Acta 1350: 128-132); GlpQ antigen of Borrelia hermsii (Schwan et al. (1996) J Clin. Microbiol. 34: 2483-2492); cholera protective antigen (CPA) (Sciortino (1996) J. Diarrhoeal Dis. Res. 14: 16-26); a 190-kDa protein antigen of Streptococcus mutans (Senpuku et al. (1996) Oral Microbiol. Immunol. 11: 121-128); Anthrax toxin protective antigen (PA) (Sharma et al. (1996) Protein Expr. Purif. 7: 33-38); Clostridium perfringens antigens and toxoid (Strom et al. (1995) Br. J. Rheumatol. 34: 1095-1096); the SEF14 fimbrial antigen of Salmonella enteritidis (Thorns et al. (1996) Microb. Pathog. 20: 235-246); the Yersinia pestis capsular antigen (F I antigen) (Titball et al. (1997) Infect. Immun. 65: 1926- 1930); a 35-kilodalton protein of Mycobacterium leprae (Triccas et al. (1996) Infect. Immun. 64: 5171-5177); the major outer membrane protein, CD, extracted from Moraxella (Branhamella) catarrhalis (Yang et al. (1997) FEMS Immunol. Med. Microbiol. 17: 187-199); pH6 antigen (PsaA protein) of Yersinia pestis (Zav'yalov et al. (1996) FEMS Immunol. Med. Microbiol. 14: 53-57); a major surface glycoprotein, gp63, of Leishmania major (Xu and Liew (1994) Vaccine 12: 1534-1536; Xu and Liew (1995) Immunology 84: 173-176); mycobacterial heat shock protein 65, mycobacterial antigen (Mycobacterium leprae hsp65) (Lowrie et al. (1994) Vaccine 12: 1537-1540; Ragno et al. (1997) Arthritis Rheum. 40: 277-283; Silva (1995) Braz. J Med. Biol. Res. 28: 843-851); Mycobacterium tuberculosis antigen 85 (Ag85) (Huygen et al. (1996) Nat. Med. 2: 893-898); the 45/47 kDa antigen complex (APA) of Mycobacterium tuberculosis, M. bovis and BCG (Horn et al. (1996) J Immunol. Methods 197: 151-159); the mycobacterial antigen, 65-kDa heat shock protein, hsp65 (Tascon et al. (1996) Nat. Med. 2: 888-892); the mycobacterial antigens MPB64, MPB70, MPB57 and alpha antigen (Yamada et al. (1995) Kekkaku 70: 63 9-644); the M. tuberculosis 3 8 kDa protein (Vordenneier et al. (1995) Vaccine 13: 1576-1582); the MPT63, MPT64 and MPT- 59 antigens from Mycobacterium tuberculosis (Manca et al. (1997) Infect. Immun. 65: 16- 23; Oettinger et al. (1997) Scand. J Immunol. 45: 499-503; Wilcke et al. (1996) Tuber. Lung Dis. 77: 250-256); the 35-kilodalton protein of Mycobacterium leprae (Triccas et al. (1996) Infect. Immun. 64: 5171-5177); the ESAT-6 antigen of virulent mycobacteria (Brandt et al. (1996) J Immunol. 157: 3527-3533; Pollock and Andersen (1997) J Infect. Dis. 175: 1251- 1254);

A~vcobacterium tuberculosis 16-kDa antigen (Hspl6.3) (Chang et al. (1996) J Biol. Chem. 271: 7218-7223); and the 18-kilodalton protein of Mycobacterium leprae (Baumgart et al. (1996) Infect. Immun. 64: 2274-228 1).

Viral Pathogens

The methods of the invention are also useful for obtaining recombinant nucleic acids and polypeptides that have enhanced ability to induce an immune response against viral pathogens. While the bacterial recombinants described above are typically administered in polypeptide form, recombinants that confer viral protection are preferably administered in nucleic acid form, as genetic vaccines.

One illustrative example is the Hantaan virus. Glycoproteins of this virus typically accumulate at the membranes of the Golgi apparatus of infected cells. This poor expression of the glycoprotein prevents the development of efficient genetic vaccines against these viruses. The methods of the invention solve this problem by performing stochastic (e.g. polynucleotide shuffling & interrupted synthesis) and non-stochastic polynucleotide reassembly on nucleic acids that encode the glycoproteins and identifying those recombinants that exhibit enhanced expression in a host cell, and/or for improved immunogenicity when administered as a genetic vaccine. A convenient screening method for these methods is to express the experimentally generated polynucleotides as fusion proteins to PIG, which results in display of the polypeptides on the surface of the host cell (Whitehorn et al. (1995) Biotechnology (N Y) 13:1215-9). Fluorescence-activated cell sorting is then used to sort and recover those cells that express an increased amount of the antigenic polypeptide on the cell surface. This preliminary screen can be followed by immunogenicity tests in mammals, such as mice. Finally, in preferred embodiments, those recombinant nucleic acids are tested as genetic vaccines for their ability to protect a test animal against challenge by the virus.

The flaviviruses are another example of a viral pathogen for which the methods of the invention are useful for obtaining a experimentally generated polypeptide or genetic vaccine that is effective against a viral pathogen. The flaviviruses consist of three clusters of antigenically related viruses: Dengue 1-4 (62-77% identity), Japanese, St. Louis and Murray Valley encephalitis viruses (75-82% identity), and the tick-borne encephalitis viruses (77-96% identity). Dengue virus can induce protective antibodies against SLE and Yellow fever (40-50% identity), but few efficient vaccines are available. To obtain genetic vaccines and

experimentally generated polypeptides that exhibit enhanced cross-reactivity and immunogenicity, the polynucleotides that encode envelope proteins of related viruses are subjected to stochastic (e.g. polynucleotide shuffling & interrupted synthesis) and non-stochastic polynucleotide reassembly. The resulting experimentally generated polynucleotides can be tested, either as genetic vaccines or by using the expressed polypeptides, for ability to induce a broadly reacting neutralizing antibody response. Finally, those clones that are favorable in the preliminary screens can be tested for ability to protect a test animal against viral challenge.

Viral antigens that can be evolved by stochastic (e.g. polynucleotide shuffling & interrupted synthesis) and non-stochastic polynucleotide reassembly for improved activity as vaccines include, but are not limited to, influenza A virus N2 neuraminidase (Kilbourne et al. (1995) Vaccine 13: 1799-1803); Dengue virus envelope (E) and premembrane (prM) antigens (Feighny et al. (1994) Am. J Trop. Med. Hyg. 50: 322-328; Putnak et al. (1996) Am. J Trop. Med. Hyg. 5 5: 5 04-10); HIV antigens Gag, Pol, Vif and Nef (Vogt et al. (1995) Vaccine 13: 202-208); HIV antigens gp 120 and gp 160 (Achour et al. (1995) Cell. Mol. Biol. 41: 395-400; Hone et al. (1994) Dev. Biol. Stand. 82: 159-162); gp4l epitope of human immunodeficiency virus (Eckhart et al. (1996) J Gen. Virol. 77: 2001- 2008); rotavirus antigen VP4 (Mattion et al. (1995) J Virol. 69: 5132-5137); the rotavirus protein VP7 or VP7sc (Emslie et al. (1995) J Virol. 69: 1747-1754; Xu et al. (1995) J Gen. Virol. 76: 1971-1980); herpes simplex virus (HSV) glycoproteins gB, gC, gD, gE, gG, gH, and gl (Fleck et al. (1994) Med. Microbiol. Immunol. (Berl) 183: 87-94 [Mattion, 1995]; Ghiasi et al. (1995) Invest. Ophthalmol. Vis. Sci. 36: 1352-1360; McLean et al. (1994) J Infect. Dis. 170: 1100-1109); immediate-early protein ICP47 of herpes simplex virus- type 1 (HSV-1) (Banks et al. (1994) Virology 200: 23 6-245); immediate-early (IE) proteins ICP27, ICPO, and ICP4 of herpes simplex virus (Manickan et al. (1995) J Virol. 69: 4711-4716); influenza virus nucleoprotein and hemagglutinin (Deck et al. (1997) Vaccine 15: 71- 78; Fu et al. (1997) J Virol. 71: 2715-272 1); B 19 parvovirus capsid proteins VP1 (Kawase et al. (1995) Virology 211: 359-366) or VP2 (Brown et al. (1994) Virology 198: 477- 488); Hepatitis B virus core and e antigen (Schodel et al. (1996) Intervirology 39:104-106); hepatitis B surface antigen (Shiau and Murray (1997) J. Med. Virol. 51: 159-166); hepatitis B surface antigen fused to the core antigen of the virus (Id.); Hepatitis B virus core-preS2 particles (Nemeckova et al. (1996) Acta Virol. 40: 273-279); HBV preS2-S protein (Kutinova et al. (1996) Vaccine 14:

1045-1052); VZV glycoprotein I (Kutinova et al. (1996) Vaccine 14: 1045-1052); rabies virus glycoproteins (Xiang et al. (1994) Virology 199: 132-140; Xuan et al. (1995) Virus Res. 36: 151-161) or ribonucleocapsid (Hooper eta/. (1994) Proc. Nat'l. Acad. Sci. USA 91: 10908-10912); human cytomegalovirus (HCMV) glycoprotein B (LTL55) (Britt et al. (1995) J Infect. Dis. 171: 18-25); the hepatitis C virus (HCV) nucleocapsid protein in a secreted or a nonsecreted form, or as a fusion protein with the middle (pre-S2 and S) or major (S) surface antigens of hepatitis B virus (HBV) (Inchauspe et al. (1997) DNA Cell Biol. 16: 185-195; Major et al. (1995) J Virol. 69: 5798-5805); the hepatitis C virus antigens: the core protein (pC); E1 (pE1) and E2 (pE2) alone or as fusion proteins (Saito et al. (1997) Gastroenterology 112: 1321-1330); the gene encoding respiratory syncytial virus fusion protein (PFP-2) (Falsey and Walsh (1996) Vaccine 14: 1214-1218; Piedra et al. (1996) Pediatr. Infect. Dis. J. 15: 23-3 1); the VP6 and VP7 genes of rotaviruses (Choi et al. (1997) Virology 232: 129-13 8; Jin et al. (1996) Arch. Virol. 141: 2057-2076); the E 1, E2, E3, E4, E5, E6 and E7 proteins of human papillornavirus (Brown et al. (1994) Virology 201: 46-54; Dillner et al. (1995) Cancer Detect. Prev. 19: 3 81- 393; Krul et al. (1996) Cancer Immunol. Immunother. 43: 44-48; Nakagawa et al. (1997) J Infect. Dis. 175: 927-93 1); a human T-lymphotropic virus type I gag protein (Porter et al. (1995) J Med Virol. 45: 469-474); Epstein-Barr virus (EBV) gp340 (Mackett et al. (1996) J Med. Virol. 50: 263-271); the Epstein-Barr virus (EBV) latent membrane protein LMP2 (Lee et al. (1996) Eur. J Immunol. 26: 1875-1883); Epstein-Barr virus nuclear antigens 1 and 2 (Chen and Cooper (1996) J Virol. 70: 4849-4853; Khanna et al. (1995) Virology 214: 633-637); the measles virus nucleoprotein (N) (Fooks et al. (1995) Virology 210: 456-465); and cytomegalovirus glycoprotein gB (Marshall et al. (1994) J Med. Virol. 43: 77-83) or glycoprotein gH (Rasmussen et al. (1994) J Infect. Dis. 170: 673-677).

Inflammatory And Autoimmune Diseases

Autoimmune diseases are characterized by immune response that attacks tissues or cells of ones own body, or pathogen-specific immune responses that also are harmful for ones own tissues or cells, or non-specific immune activation which is harmful for ones own tissues or cells. Examples of autoinimune diseases include, but are not limited to, rheumatoid arthritis, SLE, diabetes mellitus, myasthenia gravis, reactive arthritis, ankylosing spondylitis, and multiple sclerosis. These and other inflammatory conditions, including IBD, psoriasis, pancreatitis, and various immunodeficiencies, can be treated using genetic vaccines that

include vectors and other components obtained using the methods of the invention (e.g. using antigens that are optimized using the methods of the invention).

These conditions are often characterized by an accumulation of inflammatory cells, such as lymphocytes, macrophages, and neutrophils, at the sites of inflammation. Altered cytokine production levels are often observed, with increased levels of cytokine production. Several autoimmune diseases, including diabetes and rheumatoid arthritis, are linked to certain MHC haplotypes. Other autoimmune-type disorders, such as reactive arthritis, have been shown to be triggered by bacteria such as Yersinia and Shigella, and evidence suggests that several other autoimmune diseases, such as diabetes, multiple sclerosis, rheumatoid arthritis, may also be initiated by viral or bacterial infections in genetically susceptible individuals.

Current strategies of treatment generally include anti-inflammatory drugs, such as NSAID or cyclosporin, and antiproliferative drugs, such as methotrexate. These therapies are non-specific, so a need exists for therapies having greater specificity, and for means to direct the immune responses towards the direction that inhibits the autoimmune process.

The present invention provides several strategies by which these needs can be fulfilled. First, the invention provides methods of obtaining vaccines which exhibit improved delivery of tolerogenic antigens (e.g. methods of obtaining antigens having greater tolerogenicity and/or have improved antigenicity), antigens which have improved antigenicity, genetic vaccine-mediated tolerance, and modulation of the immune response by inclusion of appropriate accessory molecules. In a preferred embodiment, the vaccines (e.g. optimized antigens) prepared according to the invention exhibit improved induction of tolerance by oral delivery.

Oral tolerance is characterized by induction of immunological tolerance after oral administration of large quantities of antigen (Chen et al. (1995) Science 265: 123 7- 1240; Haq et al. (1995) Science 268: 714-716). In animal models, this approach has proven to be a very promising approach to treat autoimmune diseases, and clinical trials are in progress to address the efficacy of this approach in the treatment of human autoimmune diseases, such as rheumatoid arthritis and multiple sclerosis (Chen et al. (1994) Science 265:123 7-40; Whitacre et al. (1996) Clin. Immunol. Immunopathol. 80: S31-9; Hohol et al. (1996) Ann. N.Y Acad Sci. 778:243-50). It has also been suggested that induction of oral tolerance against viruses used in gene therapy might reduce the immunogenicity of gene therapy vectors.

However, the amounts of antigen required for induction of oral tolerance are very high and improved methods for oral delivery of antigenic proteins would significantly improve the efficacy of induction of oral tolerance.

Expression library immunization (Barry et al. (1995) Nature 3 77: 632) is a particularly useful method of screening for optimal antigens for use in genetic vaccines. For example, to identify autoantigens present in Yersinia, Shigella, and the like, one can screen for induction of T cell responses in HLA-B27 positive individuals. Complexes that include epitopes of bacterial antigens and MHC molecules associated with autoimmune diseases, e.g., HLA-B27 in association with Yersinia antigens can be used in the prevention of reactive arthritis and ankylosing spondylitis in HLA-B27 positive individuals.

Treatment of autoimmune and inflammatory conditions can involve not only administration of tolerogenic antigens, but also the use of a combination of cytokines, costimulatory molecules, and the like. Such cocktails are formulated for induction of a favorable immune response, typically induction of autoantigen-specific tolerance. Cocktails can also include, for example, CD1, which is crucially involved in recognition of self antigens by a subset of T cells (Porcelli (1995) Adv. Immunol. 5 9: 1). Genetic vaccine vectors and cocktails that skew immune responses towards the T_H2 are often used in treating autoimmune and inflammatory conditions, both with antigen-specific and antigen non-specific vectors.

Screening of genetic vaccines and accessory molecules (e.g. and optimized antigens) can be done in animal models which are known to those of skill in the art. Examples of suitable models for various conditions include collagen induced arthritis, the NFS/sld mouse model of human Sjogren's syndrome; a 120 kD organ-specific autoantigen recently identified as an analog of human cytoskeletal. protein α-fodrin (Haneji et al. (1997) Science 276: 604), the New Zealand Black/White F1 hybrid mouse model of human SLE, NOD mice, a mouse model of human diabetes mellitus, fas/fas ligand mutant mice, which spontaneously develop autoimmune and lymphoproliferative disorders (Watanabe-Fukunaga et al. (1992) Nature 356: 314), and experimental autoimmune encephalomyelitis (EAE), in which myelin basic protein induces a disease that resembles human multiple sclerosis.

Autoantigens (that can be experimentally evolved (e.g. by polynucleotide reassembly &/or polynucleotide site-saturation mutagenesis) according to the methods of the invention) that are useful in genetic vaccines for treating multiple sclerosis include, but are not limited

to, myelin basic protein (Stinissen et al. (1996) J Neurosci. Res. 45: 500-511) or a fusion protein of myelin basic protein and proteolipid protein in multiple sclerosis (Elliott et al. (1996) J Clin. Invest. 98: 1602-1612), proteolipid protein (PLP) (Rosener et al. (1997) J Neuroimmunol. 75: 28-34), 2',3'-cyclic nucleotide 3'- phosphodiesterase (CNPase) (Rosener et al. (1997) J Neuroimmunol. 75: 28- 34), the Epstein Barr virus nuclear antigen-1 (EBNA-1) in multiple sclerosis (Vaughan et al. (1996) J Neuroimmunol. 69: 95-102), HSP70 in multiple sclerosis (Salvetti et al. (1996) J Neuroimmunol. 65: 143-53; Feldmann et al. (1996) Cell 85: 307).

Target antigens that, after reassembly (optionally in combination with other directed evolution methods described herein) according to the methods of the invention, can be used to treat scleroderma, systemic sclerosis, and systemic lupus erythematosus include, for example, (-2-GPI, 50 kDa glycoprotein (Blank et al. (1994) J Autoimmun. 7: 441-455), Ku (p70/p80) autoantigen, or its 80-kd subunit protein (Hong et al. (1994) Invest. Ophthalmol. Vis. Sei. 35: 4023-4030; Wang et al. (1994) J Cell Sci. 107: 3223-3233), the nuclear autoantigens La (SS-B) and Ro (SS-A) (Huang et al. (1997) J Clin. Immunol. 17: 212-219; lgarashi et al. (1995) Autoimmunity 22: 33-42; Keech et al. (1996) Clin. Exp. Immunol. 104: 255-263; Manoussakis et al. (1995) J Autoimmun. 8: 959-969; Topfer et al. (1995) Proc. Nat'l. Acad. Sci. USA 92: 875-879), proteasome (-type subunit C9 (Feist et al. (1996) J Exp. Med. 184: 1313-1318), Scleroderma antigens Rpp 30, Rpp 38 or Scl-70 (Eder et al. (1997) Proc. Nat'l. Acad. Sci. USA 94: 1101-1106; Hietarinta et al. (1994) Br. J Rheumatol. 33: 323-326), the centrosome autoantigen PCM-1 (Bao et al. (1995) Autoimmunitv 22: 219-228), polymyositis-scleroderma autoantigen (PM-Scl) (Kho et al. (1997) J Biol. Chem. 272: 13426-1343 1), scleroderma (and other systemic autoimmune disease) autoantigen CENP-A (Muro et al. (1996) Clin. Immunol. Immunopathol. 78: 86-89), U5, a small nuclear ribonucleoprotein (snRNP) (Okano et al. (1996) Clin. Immunol. Immunopathol. 81: 41-47), the I 00-kd protein of PM-Scl autoantigen (Ge et al. (1996) Arthritis Rheum. 39: 1588-1595), the nucleolar U3- and Th(7-2) ribonucleoproteins (Verheijen et al. (1994) J. Immunol. Methods 169: 173-182), the ribosomal protein L7 (Neu et al. (1995) Clin. Exp. Immunol. 100: 198-204), hPop 1 (Lygerou et al. (1996) EMBO J. 15: 5936-5948), and a 36-kd protein from nuclear matrix antigen (Deng et al. (1996) Arthritis Rheum. 39: 1300-1307).

Hepatic autoimmune disorders can also be treated using improved recombinant antigens that are prepared according to the methods described herein. Among the antigens

that are useful in such treatments are the cytochromes P450 and UDP- glucuronosyl-transferases (Obermayer-Straub and Manns (1996) Baillieres Clin. Gastroenterol. 10: 501-532), the cytochromes P450 2C9 and P450 1A2 (Bourdi et al. (1996) Chem. Res. Toxicol. 9: 1159-1166; Clemente et al. (1997) J Clin. Endocrinol. Metab. 82: 1353-1361), LC-1 antigen (Klein et al. (1996) J Pediatr. Gastroenterol. Nutr. 23: 461-465), and a 230-kDa Golgi-associated protein (Funaki et al. (1996) Cell Struct. Funct. 21: 63-72).

For treatment of autoimmune disorders of the skin, useful antigens include, but are not limited to, the 450 kD human epidermal autoantigen (Fujiwara et al. (1996) J Invest. Dermatol. 106: 1125-1130), the 230 kD and 180 kD bullous pemphigoid antigens (Hashimoto (1995) Keio J Med. 44: 115 -123; Murakami et al. (1996) J Dermatol. Sci. 13: 112-117), pemphigus foliaceus antigen (desmoglein 1), pemphigus vulgaris antigen (desmoglein 3), BPAg2, BPAg1, and type VII collagen (Batteux et al. (1997) J Clin. Immunol. 17: 228-233; Hashimoto et al. (1996) J Dermatol. Sci. 12: 10- 17), a 168-kDa mucosal antigen in a subset of patients with cicatricial pemphigoid (Ghohestani et al. (1996) J Invest. Dermatol. 107: 136-139), and a 218-kd nuclear protein (218-kd Mi-2) (Seelig et al. (1995) Arthritis Rheum. 38: 1389-1399).

The methods of the invention are also useful for obtaining improved antigens for treating insulin dependent diabetes mellitus, using one or more of antigens which include, but are not limited to, insulin, proinsulin, GAD65 and GAD67, heat-shock protein 65 (hsp65), and islet-cell antigen 69 (ICA69) (French et al. (1997) Diabetes 46: 34-39; Roep (1996) Diabetes 45: 1147-1156; Schloot et al. (1997) Diabetologia 40: 332-338), viral proteins homologous to GAD65 (Jones and Crosby (1996) Diabetologia 39: 1318-1324), islet cell antigen-related protein-tyrosine phosphatase (PTP) (Cui et al. (1996) J Biol. Chem. 271: 24817-24823), GM2-1 ganglioside (Cavallo et al. (1996) J Endocrinol. 150: 113-120; Dotta et al. (1996) Diabetes 45: 1193 -1196), glutarnic acid decarboxylase (GAD) (Nepom (1995) Curr. Opin. Immunol. 7: 825-830; Panina-Bordignon et al. (1995) J Exp. Med. 181: 1923-1927), an islet cell antigen (ICA69) (Karges et al. (1997) Biochim. Biophys. Acta 1360: 97-101; Roep et al. (1996) Eur. J Immunol. 26: 1285-1289), Tep69, the single T cell epitope recognized by T cells from diabetes patients (Karges et al. (1997) Biochim. Biopkys. Acta 1360: 97-101), ICA 512, an autoantigen of type I diabetes (Solimena et al. (1996) EMBOJ. 15: 2102-2114), an islet-cell protein tyrosine phosphatase and the 37- kDa autoantigen derived from it in type I diabetes (including IA-2, IA-2) (La Gasse et al. (1997) Mol. Med. 3:

163-173), the 64 kDa protein from In- 111 cells or human thyroid follicular cells that is immunoprecipitated with sera from patients with islet cell surface antibodies (ICSA) (Igawa et al. (1996) Endocr. J. 43: 299-306), phogrin, a homologue of the human transmembrane protein tyrosine phosphatase, an autoantigen of type I diabetes (Kawasaki et al. (1996) Biochem. Biophys. Res. Commun. 227: 440-447), the 40 kDa and 37 kDa tryptic fragments and their precursors IA-2 and IA-2 in IDDM (Lampasona et al. (1996) J Immunol. 157: 2707-2711; Notkins et al. (1996) J A utoimmun. 9: 677-682), insulin or a cholera toxoid-insulin conjugate (Bergerot et al. (1997) Proc. Nat'l. Acad. Sci. USA 94: 4610-4614), carboxypeptidase H, the human homologue of gp330, which is a renal epithelial glycoprotein involved in inducing Heymann nephritis in rats, and the 38- kD islet mitochondrial autoantigen (Arden et al. (1996) J Clin. Invest. 97: 551- 561.

Rheumatoid arthritis is another condition that is treatable using optimized antigens prepared according to the present invention. Useful antigens for rheumatoid arthritis treatment include, but are not limited to, the 45 kDa DEK nuclear antigen, in particular onset juvenile rheumatoid arthritis and iridocyclitis (Murray et al. (1997) J Rheumatol. 24: 560-567), human cartilage glycoprotein-39, an autoantigen in rheumatoid arthritis (Verheijden et al. (1997) Arthritis Rheum. 40: 1115-1125), a 68k autoantigen in rheumatoid arthritis (Blass et al. (1997) Ann. Rheum. Dis. 56: 317-322), collagen (Rosloniec et al. (1995) J Immunol. 155: 4504-4511), collagen type II (Cook et al. (1996) Arthritis Rheum. 39: 1720-1727; Trentham (1996) Ann. N. Y. Acad. Sci. 778: 306-314), cartilage link protein (Guerassimov et al. (1997) J Rheumatol. 24: 95 9-964), ezrin, radixin and moesin, which are auto-immune antigens in rheumatoid arthritis (Wagatsuma et al. (1996) Mol. Immunol. 33: 1171-1176), and mycobacterial heat shock protein 65 (Ragno et al. (1997) Arthritis Rheum. 40: 277-283).

Also among the conditions for which one can obtain an improved antigen suitable for treatment are autoimmune thyroid disorders. Antigens that are useful for these applications include, for example, thyroid peroxidase and the thyroid stimulating hormone receptor (Tandon and Weetman (1994) J R. Coll. Physicians Lond. 28: 10- 18), thyroid peroxidase from human Graves' thyroid tissue (Gardas et al. (1997) Biochem. Biophys. Res. Commun. 234: 366-370; Zimmer et al. (1997) Histochem. Cell. Biol. 107: 115-120), a 64-kDa antigen associated with thyroid-associated ophthalmopathy (Zhang et al. (1996) Clin. Immunol. Immunopathol. 80: 23 6-244), the human TSH receptor (Nicholson et al. (1996) J Mol. Endocrinol. 16: 159-170), and the 64 kDa protein from In- 111 cells or human thyroid

follicular cells that is immunoprecipitated with sera from patients with islet cell surface antibodies (ICSA) (Igawa et al. (1996) Endocr. J. 43: 299-306).

Other conditions and associated antigens include, but are not limited to, Sjogren's syndrome (-fodrin; Haneji et al. (1997) Science 276: 604-607), myastenia gravis (the human M2 acetylcholine receptor or fragments thereof, specifically the second extracellular loop of the human M2 acetylcholine receptor; Fu et al. (1996) Clin. Immunol. Immunopathol. 78: 203-207), vitiligo (tyrosinase; Fishman et al. (1997) Cancer 79: 1461 - 1464), a 450 kD human epidermal autoantigen recognized by serum from individual with blistering skin disease, and ulcerative colitis (chromosomal proteins HMG1 and HMG2; Sobajima et al. (199 7) Clin. Exp. Immunol. 107: 135 -140).

Allergy And Asthma

The invention also provides methods of obtaining reagents that are useful for treating allergy. In one embodiment, the methods involve making a library of experimentally generated polynucleotides that encode an allergen, and screening the library to identify those experimentally generated polynucleotides that exhibit improved properties when used as immunotherapeutic reagents for treating allergy. For example, specific immunotherapy of allergy using natural antigens carries a risk of inducing anaphylaxis, which can be initiated by cross-linking of high-affinity IgE receptors on mast cells. Therefore, allergens that are not recognized by pre-existing IgE are desirable. The methods of the invention provide methods by which one can obtain such allergen variants. Another improved property of interest is induction of broader immune responses, increased safety and efficacy.

Genetic vaccine vectors and other reagents obtained using the methods of the invention can be used to treat allergies and asthma. Allergic immune responses are results of complex interactions between B cells, T cells, professional antigen- presenting cells (APC), eosinophils and mast cells. These cells take part in allergic immune responses both as modulators of the immune responses and are also involved in producing factors directly involved in initiation and maintenance of allergic responses.

Synthesis of polyclonal and allergen-specific IgE requires multiple interactions between B cells, T cells and professional antigen- presenting cells (APC).

Activation of naive, unprimed B cells is initiated when specific B cells recognize the allergen by cell surface immunoglobulin (sIg). However, costimulatory molecules expressed by activated T cells in both soluble and membrane-bound forms are necessary for

differentiation of B cells into IgE-secreting plasma cells. Activation of T helper cells requires recognition of an antigenic peptide in the context of MHC class II molecules on the plasma membrane of APC, such as monocytes, dendritic cells, Langerhans cells or primed B cells. Professional APC can efficiently capture the antigen and the peptide-MHC class II complexes are formed in a post-Golgi, proteolytic intracellular compartment and subsequently exported to the plasma membrane, where they are recognized by T cell receptor (TCR) (Monaco (1995) J Leuk. Biol. 57: 543-547). In addition, activated B cells express CD80 (B7-1) and CD86 (B7-2, B70), which are the counter receptors for CD28 and which provide a costimulatory signal for T cell activation resulting in T cell proliferation and cytokine synthesis (Bluestone (1995) Immunity 2: 555-559). Since allergen-specific T cells from atopic individuals generally belong to the T_H2 cell subset, activation of these cells also leads to production of IL-4 and IL- 13, which, together with membrane- bound costimulatory molecules expressed by activated T helper cells, direct B cell differentiation into IgE-secreting plasma cells (de Vries and Punnonen, In Cytokine Regulation of Humoral Immunity: Basic and Clinical Aspects, Ed. CM Snapper, John Wiley & Sons Ltd, West Sussex, UK, p. 195-215, 1996).

Mast cells and eosinophils are key cells in inducing allergic symptoms in target organs. Recognition of specific antigen by IgE bound to high- affinity IgE receptors on mast cells, basophils or eosinophils results in crosslinking of the receptors leading to degranulation of the cells and rapid release of mediator molecules, such as histamine, prostaglandins and leukotrienes, causing allergic symptoms.

Immunotherapy of allergic diseases currently includes hyposensibilization treatments using increasing doses of allergen injected to the patient. These treatments result skewing of immune responses towards T_H1 phenotype and increase the ratio of IgG/IgE antibodies specific for allergens. Because these patients have circulating IgE antibodies specific for the allergens, these treatments include significant risk of anaphylactic reactions.

In these reactions, free circulating allergen is recognized by IgE molecules bound to high-affinity IgE receptors on mast cells and eosinophils. Recognition of the allergen results in crosslinking of the receptors leading to release of mediators, such as histamine, prostaglandins, and leukotrienes, which cause the allergic symptoms, and occasionally anaphylactic reactions. Other problems associated with hyposensibilization include low efficacy and difficulties in producing allergen extracts reproducibly.

Genetic vaccines provide a means of circumventing the problems that have limited the usefulness of previously known hyposensibilization treatments. For example, by expressing antigens on the surface of cells, such as muscle cells, the risk of anaphylactic reactions is significantly reduced. This can be achieved by using genetic vaccine vectors that encode transmembrane forms of allergens. The allergens can also be modified in such a way that they are efficiently expressed in transmembrane forms, further reducing the risk of anaphylactic reactions. Another advantage provided by the use of genetic vaccines for hyposensibilization is that the genetic vaccines can include cytokines and accessory molecules which further direct the immune responses towards the T_H1 phenotype, thus reducing the amount of IgE antibodies produced and increasing the efficacy of the treatments. Vectors can also be evolved to induce primarily IgG and IgM responses, with little or no IgE response.

Furthermore, stochastic (e.g. polynucleotide shuffling & interrupted synthesis) and non-stochastic polynucleotide reassembly can be used to generate allergens that are not recognized by the specific IgE antibodies preexisting *in vivo*, yet are capable of inducing efficient activation of allergen-specific T cells. For example, using phage display selection, one can express experimentally evolved (e.g. by polynucleotide reassembly &/or polynucleotide site-saturation mutagenesis) allergens on phage, and only those that are not recognized by specific IgE antibodies are selected. These are further screened for their capacity to induce activation of specific T cells. An efficient T cell response is an indication that the T cell epitopes are functionally intact, although the B cell epitopes were altered, as indicated by lack of binding of specific antibodies.

In these methods, polynucleotides encoding known allergens, or homologs or fragments thereof (e.g., immunogenic peptides) are inserted into DNA vaccine vectors and used to immunize allergic and asthmatic individuals. Alternatively, the experimentally evolved (e.g. by polynucleotide reassembly &/or polynucleotide site-saturation mutagenesis) allergens are expressed in manufacturing cells, such as E. coli or yeast cells, and subsequently purified and used to treat the patients or prevent allergic disease. stochastic (e.g. polynucleotide shuffling & interrupted synthesis) and non-stochastic polynucleotide reassembly can be used to obtain antigens that activate T cells but cannot induce anaphylactic reactions. For example, a library of experimentally generated polynucleotides that encode allergen variants can be expressed in cells, such as antigen presenting cells, which are than

contacted with PBMC or T cell clones from atopic patients. Those library members that efficiently activate T_H cells from the atopic patients can be identified by assaying for T cell proliferation, or by cytokine synthesis (e.g., synthesis of IL-2, IL-4, IFN-γ. Those recombinant allergen variants that are positive in the *in vitro* tests can then be subjected to *in vivo* testing.

Examples of allergies that can be treated include, but are not limited to, allergies against house dust mite, grass pollen, birch pollen, ragweed pollen, hazel pollen, cockroach, rice, olive tree pollen, ftmgi, mustard, bee venom.

Antigens of interest include those of animals, including the mite (e.g., Dermatophagoides pteronyssinus, Dermatophagoidesfarinae, Blomia tropicalis), such as the allergens der p1 (Scobie et al. (1994) Biochem. Soc. Trans. 22: 448S; Yssel et al. (1992) J Immunol. 148: 738-745), der p2 (Chua et al. (1996) Clin. Exp. Allergy 26: 829-83 7), der p3 (Smith and Thomas (1996) Clin. Exp. Allergy 26: 571-579), der p5, der p V (Lin et al. (1994) J Allergy Clin. Immunol. 94: 989-996), der p6 (Bennett and Thomas (1996) Clin. Exp. Allergy 26: 1150- 1154), der p7 (Shen et al. (1995) Clin. Exp. Allergy 25: 416-422), der f2 (Yuuki et al. (1997) Int. Arch. Allergy Immunol. 112: 44-48), der f3 (Nishiyarna et al. (1995) FEBSLett. 377: 62-66), der f7 (Shen et al. (1995) Clin. Exp. Allergy 25: 1000-1006); Mag 3 (Fujikawa et al. (1996) Mol. Immunol. 33: 311-319). Also of interest as antigens are the house dust mite allergens Tyr p2 (Eriksson et al. (1998) Eur. J Biochem. 251: 443-447), Lep d 1 (Schmidt et al. (1995) FEBS Lett. 3 70: 11-14), and glutathione S-transferase (O'Neill et al. (1995) Immunol Lett. 48: 103- 107); the 25,589 Da, 219 amino acid polypeptide with homology with glutathione S- transferases (ONeill et al. (1994) Biochim. Biophys. Acta. 1219: 521- 528); Blo t 5 (Arruda et al. (1995) Int. Arch. Allergy Immunol. 107: 456-45 7); bee venom phospholipase A2 (Carballido et al. (1994) J Allergy Clin. Immunol. 93: 758-767; Jutel et al. (1995) J Immunol. 154: 4187-4194); bovine dermal/dander antigens BDA 11 (Rautiainen et al. (1995) J. Invest. Dermatol. 105: 660-663) and BDA20 (Mantyj arvi et al. (1996) J Allergy Clin. Immunol. 97: 1297-1303); the major horse allergen Equ c1 (Gregoire et al. (1996) J Biol. Chem. 271: 32951-32959); Jumper ant M. pilosula allergen Myr p 1 and its homologous allergenic polypeptides Myr p2 (Donovan et al. (1996) Biochem. Mol. Biol. Int. 39: 877- 885); 1-13, 14, 16 kD allergens of the mite Blomia tropicalis (Caraballo et al. (1996)J Allergy Clin. Immunol. 98: 573-579); the cockroach allergens Bla g Bd90K (Helm et al. (1996) J Allergy Clin. Immunol. 98: 172-80) and Bla g 2 (Arruda et al. (1995) J Biol.

Chem. 270: 19563-19568); the cockroach Cr-PI allergens (Wu et al. (1996) J Biol. Chem. 271: 1793 7-17943); fire ant venom allergen, Sol i 2 (Schmidt et al. (1996) J Allergy Clin. Immunol. 98: 82-88); the insect Chironomus thumini major allergen Chi t 1-9 (Kipp et al. (1996) Int. Arch. Allergy Immunol. 110: 348-353); dog allergen Can f 1 or cat allergen Fel d 1 (Ingram et al. (1995) J Allergy Clin. Immunol. 96: 449-456); albumin, derived, for example, from horse, dog or cat (Goubran Botros et al. (1996) Immunology 88: 340-347); deer allergens with the molecular mass of 22 kD, 25 kD or 60 kD (Spitzauer et al. (1997) Clin. Exp. Allergy 27: 196-200); and the 20 kd major allergen of cow (Ylonen et al. (1994) J Allergy Clin. Immunol. 93: 851-858).

Pollen and grass allergens are also useful in vaccines, particularly after optimization of the antigen by the methods of the invention. Such allergens include, for example, Hor v9 (Astwood and Hill (1996) Gene 182: 53-62, Lig v 1 (Batanero et al. (1996) Clin. Exp. Allergy 26: 1401-1410); Lol p 1 (Muller et al. (1996) Int. Arch. Allergy Immunol. 109: 352-355), Lol p II (Tamborini et al. (1995) Mol. Immunol. 32: 505- 513), Lol pVA, Lol pVB (Ong et al. (1995) Mol. Immunol. 32: 295-302), Lol p 9 (Blaher et al. (1996) J Allergy Clin. Immunol. 98: 124-132); Par J I (Costa et al. (1994) FEBS Lett. 341: 182-186; Sallusto et al. (1996) J Allergy Clin. Immunol. 97: 627-637), Par j 2.0101 (Duro et al. (1996) FEBS Lett. 399: 295-298); Bet v1 (Faber et al. (1996) J Biol. Chem. 271: 19243-19250), Bet v2 (Rihs et al. (1994) Int. Arch. Allergy Immunol. 105: 190-194); Dac g3 (Guerin-Marchand et al. (1996) Mol. Immunol. 33: 797-806); Phl p 1 (Petersen et al. (1995) J Allergy Clin. Immunol. 95: 987-994), Phl p 5 (Muller et al. (1996) Int. Arch. Allergy Immunol. 109: 352- 355), Phl p 6 (Petersen et al. (1995) Int. Arch. Allergy Immunol. 108: 55-59); Cry j I (Sone et al. (1994) Biochem. Biophys. Res. Commun. 199: 619-625), Cry j II (Namba et al. (1994) FEBS Lett. 353: 124-128); Cor a 1 (Schenk et al. (1994) Eur. J Biochem. 224: 717-722); cyn d 1 (Smith et al. (1996) J Allergy Clin. Immunol. 98: 331-343), cyn d 7 (Suphioglu et al. (1997) FEBS Lett. 402: 167-172); Pha a 1 and isoforms of Pha a 5 (Suphioglu and Singh (1995) Clin. Exp. Allergy 25: 853-865); Cha o 1 (Suzuki et al. (1996) Mol. Immunol. 33: 451-460); profilin derived, e.g, from timothy grass or birch pollen (Valenta et al. (1994) Biochem. Biopkys. Res. Commun. 199:106-118); P0149(Wuet al. (1996) Plant Mol.Biol. 32: 1037-1042); Ory s1 (Xuet al. (1995) Gene 164:255-259); and Amb a V and Amb t5 (Kim et al. (1996) Mol. Immunol. 33: 873-880; Zhu et al. (1995) J Immunol. 155: 5064- 5073).

Vaccines against food allergens can also be developed using the methods of the invention. Suitable antigens for reassembly (optionally in combination with other directed evolution methods described herein) include, for example, profilin (Rihs et al. (1994) Int. Arch. Allergy Immunol. 105: 190-194); rice allergenic cDNAs belonging to the alpha-amylase/trypsin inhibitor gene family (Alvarez et al. (1995) Biochim Biophys Acta 1251: 201-204); the main olive allergen, Ole e I (Lombardero et al. (1994) Clin Exp Allergy 24: 765-770); Sin a 1, the major allergen from mustard (Gonzalez De La Pena et al. (1996) Eur J Biochem. 237: 827-832); parvalbumin, the major allergen of salmon (Lindstrom et al. (1996) Scand. J Immunol. 44: 335-344); apple allergens, such as the major allergen Mal d 1 (Vanek-Krebitz et al. (1995) Biochem. Biophys. Res. Commun. 214: 538- 551); and peanut allergens, such as Ara h I (Burks et al. (1995) J Clin. Invest. 96: 1715- 1721).

The methods of the invention can also be used to develop recombinant antigens that are effective against allergies to fungi. Fungal allergens useful in these vaccines include, but are not limited to, the allergen, Cla h III, of Cladosporium herbarum (Zhang et al. (1995) J Immunol. 154: 710-717); the allergen Psi c 2, a fungal cyclophilin, from the basidiomycete Psilocybe cubensis (Homer et al. (1995) Int. Arch. Allergy Immunol. 107: 298-300); hsp 70 cloned from a cDNA library of Cladosporium herbarum (Zhang et al. (1996) Clin Exp Allergy 26: 88-95); the 68 kD allergen of Penicillium notatum (Shen et al. (1995) Clin. Exp. Allergy 26: 350-356); aldehyde dehydrogenase (ALDH) (Achatz et al. (1995) Mol Immunol. 32: 213-227); enolase (Achatz et al. (1995) Mol. Immunol. 32: 213- 227); YCP4 (Id.); acidic ribosomal protein P2 (Id.).

Other allergens that can be used in the methods of the invention include latex allergens, such as a major allergen (Hev b 5) from natural rubber latex (Akasawa et al. (1996) J Biol. Chem. 271: 25389-25393; Slater et al. (1996) J Biol. Chem. 271: 25394- 25399).

The invention also provides a solution to another shortcoming of genetic vaccination as a treatment for allergy and asthma. While genetic vaccination primarily induces $CD8^+$ T cell responses, induction of allergen-specific IgE responses is dependent on $CD4^+$ T cells and their help to B cells. T_H2-type cells are particularly efficient in inducing IgE synthesis because they secrete high levels of IL-4, IL-5 and IL-13, which direct Ig isotype switching to IgE synthesis. IL-5 also induces eosinophilia. The methods of the invention can be used to develop genetic vaccines that efficiently induce $CD4^+$ T cell responses, and direct differentiation of these cells towards the T_H1 phenotype.

The invention also provides methods by which the level of antigen release by a genetic vaccine vector is regulated. Regulation of the antigen dose is crucial at the onset of hyposensibilization for safety reasons. Low antigen levels are preferably used at first, with the antigen level increasing once evidence has been obtained that the antigen does not induce adverse effects in the individual. The stochastic (e.g. polynucleotide shuffling & interrupted synthesis) and non-stochastic polynucleotide reassembly methods of the invention allow generation of genetic vaccine vectors that induce expression of different (high and low) levels of antigen. For example, two or more different evolved promoters can be used for antigen expression. Alternatively, the antigen gene itself can be evolved for different levels of expression by, for example, altering codon usage. Vectors that induce different levels of antigen expression can be screened by use of specific monoclonal antibodies, and cell sorting (e.g, FACS).

Cancer

Immunotherapy has great promise for the treatment of cancer and prevention of metastasis. By inducing an immune response against cancerous cells, the body's immune system can be enlisted to reduce or eliminate cancer. (e.g. using the improved antigens obtained using the methods of the invention). Genetic vaccines prepared using the methods of the invention, as well as accessory molecules described herein, provide cancer immunotherapies of increased effectiveness compared to those that are presently available.

One approach to cancer immunotherapy is vaccination using genetic vaccines that include or encode antigens that are specific for tumor cells or by injecting the patients with purified recombinant cancer antigens. The methods of the invention can be used for (obtaining antigens that exhibit an) enhancement of immune responses against known tumor-specific antigens, and also to search for novel protective antigenic sequences. Genetic vaccines that exhibit optimized antigen expression, processing, and presentation can be obtained as described herein. The methods of the invention are also suitable for obtaining optimized cytokines, costimulatory molecules, and other accessory molecules that are effective in induction of an antitumor immune response, as well as for obtaining genetic vaccines and cocktails that include these and other components present in optimal combinations. The approach used for each particular cancer can vary. For treatment of hormone-sensitive cancers (for example, breast cancer and prostate cancer), methods of the invention can be used to obtain optimized hormone antagonists. For highly immunogenic

tumors, including melanoma, one can screen for genetic vaccine vectors (recombinant antigen) that optimally boost the immune response against the tumor.

Breast cancer, in contrast, is of relatively low immunogenicity and exhibits slow progression, so individual treatments can be designed for each patient. Prevention of metastasis is also a goal in design of genetic vaccines.

Among the tumor-specific antigens that can be used in the antigen reassembly (optionally in combination with other directed evolution methods described herein) methods of the invention are: bullous pemphigoid antigen 2, prostate mucin antigen (PMA) (Beckett and Wright (1995) Int. J Cancer 62: 703-710), tumor associated Thomsen- Friedenreich antigen (Dahlenborg et al. (1997) Int. J Cancer 70: 63-71), prostate-specific antigen (PSA) (Dannull and Belldegrun (1997) Br. J Urol. 1: 97-103), luminal epithelial antigen (LEA. 135) of breast carcinoma and bladder transitional cell carcinoma (TCC) (Jones et al. (1997) Anticancer Res. 17: 685-687), cancer-associated serum antigen (CASA) and cancer antigen 125 (CA 125) (Kierkegaard et al. (1995) Gynecol. Oncol. 59: 251-254), the epithelial glycoprotein 40 (EGP40) (Kievit et al. (1997) Int. J Cancer 71: 237-245), squamous cell carcinoma antigen (SCC) (Lozza et al. (1997) Anticancer Res. 17: 525-529), cathepsin E (Mota et al. (1997) Ant. J Pathol. 150: 1223-1229), tyrosinase in melanoma (Fishman et al. (1997) Cancer 79: 1461-1464), cell nuclear antigen (PCNA) of cerebral cavemomas (Notelet et al. (1997) Surg. Neurol. 47: 364-370), DF3/MUCl breast cancer antigen (Apostolopoulos et al. (1996) Immunol. Cell. Biol. 74: 45 7-464; Pandey et al. (1995) Cancer Res. 5 5: 4000-4003), carcinoembryonic antigen (Paone et al. (1996) J Cancer Res. Clin. Oncol. 122: 499-503; Schlom et al. (1996) Breast Cancer Res. Treat. 38: 27-39), tumor-associated antigen CA 19-9 (Tolliver and O'Brien (1997) South Med. J. 90: 89-90; Tsuruta et al. (1997) Urol. Int. 5 8: 20-24), human melanoma antigens MART- I /Melan-A27- and gplOO (Kawakami and Rosenberg (1997) Int. Rev. Immunol. 14: 173-192; Zajac et al. (1997) Int. J Cancer 71: 491-496), the T and Tn pancarcinoma (CA) glycopeptide epitopes (Springer (1995) Crit. Rev. Oncog. 6: 57-85), a 35 kD tumor-associated autoantigen in papillary thyroid carcinoma (Lucas et al. (1996) Anticancer Res. 16: 2493 -2496), KH- I adenocarcinoma antigen (Deshpande and Danishefsky (1997) Nature 387: 164-166), the A60 mycobacterial antigen (Maes et al. (1996) J Cancer Res. Clin. Oncol. 122: 296-300), heat shock proteins (HSPs) (Blachere and Srivastava (1995) Semin. Cancer Biol. 6: 349-355), and MAGE, tyrosinase, melan-A and gp75 and mutant oncogene products (e.g., p53, ras, and HER-2/neu (Bueler and

Mulligan (1996) Mol. Med. 2: 545-555; Lewis and Houghton (1995) Semin. Cancer Biol. 6: 321-327; Theobald et al. (1995) Proc. Nat'l. Acad. Sci. USA 92: 11993-11997).

Parasites

Antigens from parasites can also be optimized by the methods of the invention. These include, but are not limited to, the schistosome gut- associated antigens CAA (circulating anodic antigen) and CCA (circulating cathodic antigen) in Schistosoma mansoni, S. haematobium or S. japonicum (Deelder et al. (1996) Parasitology 112: 21-35); a multiple antigen peptide (MAP) composed of two distinct protective antigens derived from the parasite Schistosoma mansoni (Ferru et al. (1997) Parasite Immunol. 19: 1 -11); Leishmania parasite surface molecules (Lezama-Davila (1997) Arch. Med Res. 28: 47-53); third-stage larval (L3) antigens of L. loa (Akue et al. (1997) J Infect. Dis. 175: 158-63); the genes, Tams 1-1 and Tams 1-2, encoding the 30-and 32-kDa major merozoite surface antigens of Theileria annulata (Ta) (d'Oliveira et al. (1996) Gene 172: 33-39); Plasmodium falciparum merozoite surface antigen 1 or 2 (al-Yaman et al. (1995) Trans. R. Soc. Trop. Med. Hyg. 89: 555-559; Beck et al. (1997) J Infect. Dis. 175: 921-926; Rzepczyk et al. (1997) Infect. Immun. 65: 1098-1100); circumsporozoite (CS) protein- based B-epitopes from Plasmodium berghei, (PPPPNPND)2 and Plasmodium yoelii, (QGPGAP)3QG, along with a P. berghei T-helper epitope KQIRDSITEEWS (Reed et al. (1997) Vaccine 15: 482-488); NYVAC-Pf7 encoded Plasmodium falciparum antigens derived from the sporozoite (circumsporozoite protein and sporozoite surface protein 2), liver (liver stage antigen 1), blood (merozoite surface protein 1, serine repeat antigen, and apical membrane antigen 1), and sexual (25-kDa sexual-stage antigen) stages of the parasite life cycle were inserted into a single NYVAC genome to generate NYVAC-Pf7 (Tine et al. (1996) Infect. Immun. 64: 3833-3844); Plasmodium falciparum antigen Pfs230 (Williamson et al. (1996) Mol. Biochem. Parasitol. 78: 161-169); Plasmodium falciparum apical membrane antigen (AMA- 1) (Lal et al. (1996) Infect. Immun. 64: 1054-1059); Plasmodium falciparum proteins Pfs28 and Pfs25 (Duffy and Kaslow (1997) Infect. Immun. 65: 1109-1113); Plasmodium falciparum merozoite surface protein, MSP1 (Hui et al. (1996) Infect. Immun. 64: 1502- 1509); the malaria antigen Pf332 (Ahlborg et al. (1996) Immunology 88: 630-635); Plasmodium falciparum erythrocyte membrane protein I (Baruch et al. (1995) Proc. Nat'l. Acad. Sci. USA 93: 3497-3502; Baruch et al. (1995) Cell 82: 77-87); Plasmodium falciparum merozoite surface antigen, PfMSP-1 (Egan et al. (1996) J Infect. Dis. 173: 765- 769); Plasmodiumfalciparum antigens SERA, EBA- 175, RAP1 and

RAP2 (Riley (1997) J Pharm. Pharmacol. 49: 21-27); Schistosoma japonicum paramyosin (Sj97) or fragments thereof (Yang et al. (1995) Biochem. Biophys. Res. Commun. 212: 1029-1039); and Hsp70 in parasites (Maresca and Kobayashi (1994) Experientia 50: 1067-1074).

Contraception

Genetic vaccines that contain optimized antigens obtained by the methods of the invention are also useful for contraception. For example, genetic vaccines can be obtained that encode sperm cell specific antigens, and thus induce anti- sperm immune responses. Vaccination can be achieved by, for example, administration of recombinant bacterial strains, e.g. Salmonella and the like, which express sperm antigen, as well as by induction of neutralizing anti-hCG antibodies by vaccination by DNA vaccines encoding human chorionic gonadotropin (hCG), or a fragment thereof.

Sperm antigens which can be used in the genetic vaccines include, for example, lactate dehydrogenase (LDH-C4), galactosyltransferase (GT), SP-10, rabbit sperm autoantigen (RSA), guinea pig (g)PH-20, cleavage signal protein (CS-1), HSA-63, human (h)PH-20, and AgX-1 (Zhu and Naz (1994) Arch. Androl. 33: 141-144), the synthetic spenn peptide, P10G (O'Rand et al. (1993) J Reprod. Immunol. 25: 89-102), the 135kD, 95kD, 65kD, 47kD, 41 kD and 23kD proteins of sperm, and the FA-1 antigen (Naz et al. (1995) Arch. Androl. 35: 225-23 1), and the 35 kD fragment of cytokeratin 1 (Lucas et al. (1996) Anticancer Res. 16: 2493-2496).

The methods of the invention can also be used to obtain genetic vaccines that are expressed specifically in testis. For example, polynucleotide sequences that direct expression of genes that are specific to testis can be used (e.g., fertilization antigen-1 and the like). In addition to sperm antigens, antigens expressed on oocytes or hormones regulating reproduction may be useful targets of contraceptive vaccines. For example, genetic vaccines can be used to generate antibodies against gonadotropin releasing hormone (GnRH) or zona pellucida proteins (Miller et al. (1997) Vaccine 15:185 8-1862). Vaccinations using these molecules have been shown to be efficacious in animal models (Miller et al. (1997) Vaccine 15:1858-1862). Another example of a useful component of a genetic contraceptive vaccine is the ovarian zona pellucida glycoprotein ZP3 (Tung et al. (1994) Reprod Fertil. Dev. 6:349-355).

Malarial Antigens And Vaccines

The present invention generally relates to the Plasmodium falciparum erythrocyte membrane protein 1 ("PfEMP1"), nucleic acids which encode PfEMP1, and antibodies which specifically recognize PfEMP1. The polypeptides, antibodies and nucleic acids are useful in a variety of applications including therapeutic, prophylactic, including vaccination, diagnostic and screening applications.

The data described herein, indicates that PfEMP1 is responsible for both antigenic variation and receptor properties on PE, both of which are central to the special virulence and pathology of P. falciparum. The central role of PfEMP1 in P. falciparum biology, as the malarial adherence receptor for host proteins on microvascular endothelium, as described herein, indicates its usefulness in a malaria vaccine, in modelling prophylactic drugs, and also as a target for therapeutics to reverse PE adherence in acute cerebral malaria (Howard and Gilladoga, 1989).

Malarial Polypeptides

Soluble PfEMP1 has been reported to bind to CD36, TSP and ICAM-1, and tryptic fragments of PfEMP1 cleaved from the PE surface have been shown to bind to TSP or CD36 (Baruch, et al., Molecular Parasitology Meeting at Woods Hole, Sept 18- 22, 1994). Accordingly, in one aspect, the present invention provides substantially pure PfEMP1 polypeptides, analogs or biologically active fragments thereof.

The terms "substantially pure" or "isolated" refer, interchangeably, to proteins, polypeptides and nucleic acids which are separated from proteins or other contaminants with which they are naturally associated. A protein or polypeptide is considered substantially pure when that protein makes up greater than about 50% of the total protein content of the composition containing that protein, and typically, greater than about 60% of the total protein content. More typically, a substantially pure protein will make up from about 75 to about 90% of the total protein. Preferably, the protein will make up greater than about 90%, and more preferably, greater than about 95% of the total protein in the composition.

The term "biologically active fragment" as used herein, refers to portions of the proteins or polypeptides, e.g., a PfEMP1 derived polypeptide, which portions possess a particular biological activity, e.g., one or more activities found in a full length PfEMP1 polypeptide. For example, such biological activity may include the ability to bind a particular protein, substrate or ligand, to elicit antibodies reactive with PE, PfEMP1, the recombinant proteins or fragments thereof, to block, reverse or otherwise inhibit an interaction between

two proteins, between an enzyme and its substrate, between an epitope and an antibody, or may include a particular catalytic activity. With regard to the polypeptides of the present invention, particularly preferred polypeptides or biologically active fragments include, e.g., polypeptides that possess one or more of the biological activities described above, such as the ability to bind a ligand of PfEMP1 or inhibit the binding of PfEMP1 to one or more of its ligands, e.g., CD36, TSP, ICAM-1, VCAM-1, ELAM-1, Chondroitin sulfate or by the presence within the polypeptide fragment of antigenic determinants which permit the raising of antibodies to that fragment.

The polypeptides of the present invention may also be characterized by their immunoreactivity with antibodies raised against PfEMP1 proteins or polypeptides. In particularly preferred aspects, the polypeptides are capable of inhibiting an interaction between a PfEMP1 protein and an antibody raised against a PfEMP1 protein. Additionally or alternatively, such fragments may be specifically immunoreactive with an antibody raised against a PfEMP1 protein. Such fragments are also referred to herein as "immunologically active fragments." Generally, such biologically active fragments will be from about 5 to about 500 amino acids in length.

Typically, these peptides will be from about 20 to about 250 amino acids in length, and preferably from about 50 to about 200 amino acids in length. Generally, the length of the fragment may depend, in part, upon the application for which the particular peptide is to be used. For example, for raising antibodies, the peptides may be of a shorter length, e.g., from about 5 to about 50 amino acids in length, whereas for binding applications, the peptides may have a greater length, e.g., from about 50 to about 500 amino acids in length, preferably, from about 100 to about 250 amino acids in length, and more preferably, from about 100 to about 200 amino acids in length.

The polypeptides of the present invention may generally be prepared using recombinant or synthetic methods well known in the art. Recombinant techniques are generally described in Sambrook, et al., Molecular Cloning: A Laboratory Manual, (2nd ed.) Vols. 1-3, Cold Spring Harbor Laboratory, (1989). Techniques for the synthesis of polypeptides are generally described in Merrifield, J. Amer. Chem. Soc. 85:2149-2456 (1963), Atherton, et al., Solid Phase Peptide Synthesis: A Practical Approach, IRL Press (1989), and. Merrifield, Science 232:341-347 (1986).

In preferred aspects, the polypeptides of the present invention may be expressed by a suitable host cell that has been transfected with a nucleic acid of the invention, as described in greater detail below. Isolation and purification of the polypeptides of the present invention can be carried out by methods that are generally well known in the art. For example, the polypeptides may be purified using readily available chromatographic methods, e.g., ion exchange, hydrophobic interaction, HPLC or affinity chromatography, to achieve the desired purity. Affinity chromatography may be particularly attractive in allowing the investigator to take advantage of the specific biological activity of the desired peptide, e.g., ligand binding, presence of antigenic determinants, or the like.

Exemplary polypeptides of the present invention will generally comprise an amino acid sequence that is substantially homologous to the amino acid sequence of a PfEMP1 protein, or biologically active fragments thereof, or may include sequences that may take on a homologous conformation. In particularly preferred aspects, the polypeptides of the present invention will comprise an amino acid sequence that is substantially homologous to the amino is acid sequence shown, described &/or referenced herein (including incorporated by reference), or a biologically active fragment thereof.

By "substantially homologous" is meant an amino acid sequence which is at least about 50% homologous to the amino acid sequence of PfEMP1 or a biologically active fragment thereof, preferably at least about 90% homologous, and wore preferably at least about 95% homologous. In some aspects, substantially homologous may include a sequence that is at least 50% homologous, but that presents a homologous structure in three dimensions, i.e., includes a substantially similar surface charge or presentation of hydrophobic groups.

Examples of preferred polypeptides include polypeptides having an amino acid sequence substantially homologous to the MC PfEMP1 amino acid sequence as shown, described &/or referenced herein (including incorporated by reference), and PfEMP1 of other P. falciparum strains as shown, described &/or referenced herein (including incorporated by reference), as well as biologically active fragments of these polypeptides. Preferred peptides include those peptide fragments of PfEMP1 that are involved in the sequestration of parasitized erythrocytes. Examples of these preferred peptides include peptides which comprise an amino acid sequence which is substantially homologous to amino acids 576

through 755 of the PfEMP1 amino acid sequence shown, described &/or referenced herein (*including incorporated by reference*).

Also among the particularly preferred peptides of the present invention are those peptides and peptide fragments of PfEMP1 which are relatively conserved among the variant strains of P. falciparum or which contain regions of high homology to PfEMP1 proteins from other strains. The term "relatively conserved" generally refers to amino acid sequences that are substantially homologous to portions of the amino acid sequence shown, described &/or referenced herein (including incorporated by reference). However, also included within the definition of this term are peptides which are encoded by a nucleic acid which is a PCR product of primer probes, and particularly, universal primers, derived from the PfEMP1 nucleic acid sequence. In particular, primer is probes derived from the nucleic acid sequence shown, described &/or referenced herein (including incorporated by reference), may be used to amplify nucleic acids from other strains of P. falciparum. Particularly preferred primer sequences include the primer sequences shown in Table 1, below. Similarly, universal primer compositions, described in greater detail below and also shown in Table 1, may be used to amplify sequences that encode the peptides of the present invention.

Specific examples of relatively conserved peptides include those that are contained in a region of PfEMP1 proteins that corresponds to amino acids 576 through 755 of the amino acid sequence of MC PfEMP1, as shown, described &/or referenced herein (including incorporated by reference).

Similar regions have been specifically elucidated in a number of P. falciparum strains (as described herein). In general, these corresponding regions may be described as containing amino acid sequences that are encoded by the universal primer sequences described below. Generally, these amino acid sequences have one or more of the following general structures:

$TTIDKX_1LX_2HE$ and/or $FFWX_3WVX_4X_5ML$

where X_1 is selected from leucine or isoleucine, X_2 is selected from glutamine and asparagine, X_3 is selected from the methionine, lysine and aspartic acid, X_4 is selected from histidine, threanine and tyrosine and X_5 is selected from aspartic acid, glutamic acid and histidine. In particularly preferred aspects, the polypeptides may contain both of the above general amino acid sequences. Particularly preferred amino acid sequences will possess the conserved amino acids shown in the various fragments shown, described &/or referenced herein (including incorporated by reference). In particular, conserved amino acid sequences

of six amino acids or greater, shown, described &/or referenced herein (including incorporated by reference), may be used as epitopes for generation of antibodies that cross react with multiple P. falciparum strains.

The peptides of the invention may be free or tethered, or may include labeled groups for detection of the presence of the polypeptides. Suitable labels include radioactive, fluorescent and catalytic labeling groups that are well known in the art and that are substantially described herein, e.g., signaling enzymes, chemical reporter groups, polypeptide signals, biotin and the like. Additionally, the peptides may include modifications to the N and C-termini of the peptide, e.g., an acylated N-terminus or amidated C- terminus.

Also included within the present invention are amino acid variants of the above described polypeptides. These variants may include insertions, deletions and substitutions with other amino acids. For example, in some aspects, amino acids may be substituted with different amino acids having similar structural characteristics, e.g., net charge, hydrophobicity, or the like. For example, phenylalanine may be substituted with tyrosine, as a similarly hydrophobic residue. Glycosylation modifications, either changed, increased amounts or decreased amounts, as well as other sequence modifications are also envisioned.

In addition to the above polypeptides which consist only of naturally-occurring amino acids, peptidomimetics of the polypeptides of the present invention are also provided. Peptide analogs are commonly used in the pharmaceutical industry as non-peptide drugs with properties analogous to those of the template peptide. These types of non-peptide compound are termed "peptide mimetics" or "peptidomimetics" (Fauchere, J. (1986) Adv. Drug Res. 15:29; Veber and Freidinger (1985) TINS p.392; and Evans et al. (1987) J. Med. Chem 30:1229, and are usually developed with the aid of computerized molecular modeling. Peptide mimetics that are structurally similar to therapeutically useful peptides may be used to produce an equivalent therapeutic or prophylactic effect. Generally, peptidomimetics are structurally similar to a paradigm polypeptide (i.e., a polypeptide that has a biological or pharmacological activity), such as naturally- occurring receptor-binding polypeptide, but have one or more peptide linkages optionally replaced by a linkage selected from the group consisting of: $-CH_2NH-$, $-CH_2S-$, $-CH_2-CH_2-$, - CH=CH- (cis and trans), $-COCH_2-$, $-CH(OH)CH_2-$, and $-CH_2SO-$, by methods known in the art and further described in the following references: Spatola, A.F. in Chemistry and Biochemistry of Amino Acids, Peptides, and Proteins, B. Weinstein, eds., Marcel Dekker, New York, p. 267 (1983); Spatola, A.F.,

Vega Data (March 1983), Vol. 1, Issue 3, "Peptide Backbone Modifications" (general review); Morley, J.S., Trends Pharm Sci (1980) pp. 463-468 (general review); Hudson, D. et al., Int J Pept Prot Res (1979) 14:177-185 (- CH_2NH-, CH_2CH_2-) ; Spatola, A.F. et al., Life Sci (1986) 38:1243-1249 (-CH_2-S); Hann, M.M., J Chem Soc Perkin Trans I (1982) 307-314 (-CH-CH-, cis and trans); Almquist, R.G. et al., J Med Chem (1980) 23:1392-1398 (-$COCH_2$-); Jennings-White, C. et al., Tetrahedron Lett (1982) 23:2533 (-$COCH_2$-); Szelke, M. et al., European Appln. EP 45665 (1982) CA: 97:39405 (1982) (-$CH(OH)CH_2$-); Holladay, M.W. et al., Tetrahedxon Lett (1983) 24:4401-4404 (-$C(OH)CH_2$-); and Hruby, V.J., Life Sci (1982) 31:189-199 (-CH_2-S-)' Peptide mimetics may have significant advantages over polypeptide embodiments, including, for example: more economical production, greater chemical stability, enhanced pharmacological properties (half-life, absorption, potency, efficacy, etc.), altered specificity (e.g., a broad-spectrum of biological activities), reduced antigenicity, and others.

Labeling of peptidomimetics usually involves covalent attachment of one or more labels, directly or through a spacer (e.g., an amide group), to non-interfering position(s) on the peptidomimetic that are predicted by quantitative structure- activity data and/or molecular modeling. Such non-interfering positions generally are positions that do not form direct contacts with the molecules to which the peptidomimetic binds (e.g., CD36) to produce the therapeutic effect. Derivitization (e.g., labeling) of peptidomimetics should not substantially interfere with the desired biological or pharmacological activity of the peptidomimetic. Generally, peptidomimetics of peptides of the invention bind to their ligands (e.g., CD36) with high affinity and possess detectable biological activity (i.e., are agonistic or antagonistic to one or more ligand-mediated phenotypic changes).

Systematic substitution of one or more amino acids of a consensus sequence with a D-amino acid of the same type (e.g., D-lysine in place of L-lysine) may be used to generate more stable peptides. In addition, constrained peptides comprising a consensus sequence or a substantially identical consensus sequence variation may be generated by methods known in the art (Rizo and Gierasch (1992) Ann. Rev. Blochem. 61: 3§7; for example, by adding internal cysteine residues capable of forming intramolecular disulfide bridges which cyclize the peptide.

Polypeptides of the present invention may also be characterized by their ability to bind antibodies raised against PfEMP1, or fragments thereof. Preferably, these antibodies

recognize polypeptide domains that are homologous to the PfEMP1 proteins from a number of variants of P. falciparum. These homologous domains will generally be present throughout the family of PfEMP1 proteins. A variety of immunoassay formats may be used to select antibodies specifically immunoreactive with a particular protein or domain. For example, solid-phase ELISA immunoassays are routinely used to select monoclonal antibodies specifically immunoreactive with a protein. See Harlow and Lane (1988) Antibodies, A Laboratory Manual, Cold Spring Harbor Publications, New York, for a description of immunoassay formats and conditions that can be used to determine specific immunoreactivity. Antibodies to PfEMP1 and its fragments are discussed in greater detail, below. As used herein, the terms "polypeptide" or "peptide" are used interchangeably to refer to peptides, peptidomimetics, analogs, and the like, as described above.

The polypeptides of the present invention may be used as isolated polypeptides, or may exist as fusion proteins. A "fusion protein" generally refers to a composite protein made up of two or more separate, heterologous proteins which are normally not fused together as a single protein.

Thus, a fusion protein may comprise a fusion of two or more heterologous or homologous sequences, provided these sequences are not normally fused together. Fusion proteins will generally be made by either recombinant nucleic acid methods, i.e., as a result of transcription and translation of a gene fusion comprising a segment encoding a polypeptide comprising a PfEMP1 protein and a segment which encodes one or more heterologous proteins, or by chemical synthesis methods well known in the art.

Malarial Nucleic Acids And Cells Capable Of Exdressing Same

Also provided in the present invention are isolated nucleic acid sequences which encode the above described polypeptides and biologically active fragments. Typically, such nucleic acid sequences will comprise a segment that is substantially homologous to a portion or fragment of the nucleic acid sequence shown, described &/or referenced herein (including incorporated by reference). Preferably, the nucleic acids of the present invention will comprise at least about 15 consecutive nucleotides of the nucleic acid, more preferably, at least about 20 contiguous nucleotides, still more preferably, at least about 30 contiguous nucleotides, and still more preferably, at least about 50 contiguous nucleotides from the nucleotide sequence.

Substantial homology in the nucleic acid context means that the segments, or their complementary strands, when compared, are the same when properly aligned with the appropriate nucleotide insertions or deletions, in at least about 60% of the nucleotides, typically, at least about 70%, more typically, at least about 80%, usually, at least about 90%, and more usually, at least about 95% to 98% of the nucleotides. Alternatively, substantial homology exists when the segments will hybridize under selective hybridization conditions to a strand, or its complement, typically using a sequence of at least about 15 contiguous nucleotides derived from the PfEMP1 nucleic acid sequence. However, larger segments will usually be preferred, e.g., at least about 20 or contiguous nucleotides, more usually about 40 contiguous nucleotides, and preferably more than about 50 contiguous nucleotides. Selective hybridization exists when hybridization occurs which is more selective than total lack of specificity. See, Kanchisa, Nucleic Acid Res. 12:203-213 (1984).

Nucleic acids of the present invention include RNA, cDNA, genomic DNA, synthetic forms and mixed polymers, both sense and antisense strands. Furthermore, different alleles of each isoform are also included. The present invention also provides recombinant nucleic acids which are not otherwise naturally occurring. The nucleic acids included in the present invention will typically comprise RNA or DNA or mixed polymers. The DNA compositions will generally include a coding region which encodes a polypeptide comprising an amino acid sequence substantially homologous to the amino acid sequence of a PfEMP1 protein. More preferred are those DNA segments comprising a nucleotide sequence which encodes a CD36 binding fragment of the PfEMP1 protein.

cDNA encoding the polypeptides of the present invention, or fragments thereof, may be readily employed as a probe useful for obtaining genes which encode the PfEMP1 polypeptides of the present invention. Preparation of these probes may be carried out by generally well known methods. For example, the cDNA probes may be prepared from the amino acid sequence of the PfEMP1 protein. In particular, probes may be prepared based upon segments of the amino acid sequence which possess relatively low levels of degeneracy, i.e., few or one possible nucleic acid sequences which encode therefor.

Suitable synthetic DNA fragments may then be prepared, e.g., by the phosphoramidite method described by Beaucage and Carruthers, Tetra. Letts. 22:1859-1862 (1981). Alternatively, nucleotide sequences which are relatively conserved among the PfEMP1 coding sequences for the various P. falciparum strains may be used as suitable